The Neural Representation of Scenes in Visual Cortex

David Mark Watson

Doctor of Philosophy

University of York

Psychology

February 2016

Abstract

Recent neuroimaging studies have identified a number of regions in the human brain that respond preferentially to visual scenes. These regions are thought to underpin our ability to perceive and interact with our local visual environment. However, the precise stimulus dimensions underlying the function of scene-selective regions remain controversial. Some accounts have proposed an organisation based on relatively high-level semantic or categorical properties of the stimulus. However, other accounts have suggested that lower-level visual features of the stimulus may offer a more parsimonious explanation. This thesis presents a series of fMRI experiments employing multivariate pattern analyses (MVPA) in order to test the role of low-level visual properties in the function of scene-selective regions. The first empirical chapter presents two experiments showing that patterns of neural response to different scene categories can be predicted by a model of the visual properties of scenes (GIST). The second empirical chapter demonstrates that direct manipulations of the spatial frequency content of the image significantly influence the patterns of response, with effects often being comparable to or greater than those of scene category. The third empirical chapter demonstrates that distinct patterns of response can be found to different scene categories even when images are Fourier phase scrambled such that low-level visual features are preserved, but perception of the categories is impaired. The fourth and final empirical chapter presents an experiment using a data-driven method to select clusters of scenes objectively based on their visual properties. These visually defined clusters did not correspond to typical scene categories, but nevertheless elicited distinct patterns of neural response. Taken together, these results support the importance of low-level visual features in the functional organisation of scene-selective regions. Scene-selective responses may arise from the combined sensitivity to multiple visual features that are themselves predictive of scene content.

List of Contents

List of Tables

List of Figures

Acknowledgements

Firstly, I would like to express my deepest gratitude to my supervisors, Prof. Timothy Andrews and Dr. Tom Hartley, for their continued support over the past few years. They have afforded me a unique opportunity to gain new experiences and insights into this field of research, and never failed to provide guidance when it was needed. I would further like to thank Prof. Alex Wade, as the additional member of my thesis advisory panel, for his helpful advice throughout the PhD.

I would also like to thank the members of the York Neuroimaging Centre team, and in particular Dr. Mark Hymers and Dr. André Gouws, who have repeatedly provided me with assistance on a number of the more technical aspects of the research presented here.

Finally, I would like to thank my family for their continued moral and financial support, without which my undertaking this PhD would not have been possible.

To the memory of Lilian Sharp (1923 – 2013)

Author's Declaration

This thesis presents original work completed by the author, David Watson, under the joint supervision of Prof. Timothy Andrews and Dr. Tom Hartley, and has not been submitted to any institution other than the University of York. All sources are acknowledged as references.

The empirical work presented in this thesis has been published or is currently under review in the following peer-reviewed journals:

Watson, D. M., Hartley, T., & Andrews, T. J. (2014). Patterns of response to visual scenes are linked to the low-level properties of the image. *NeuroImage*, *99*, 402–410.

Watson, D. M., Hymers, M., Hartley, T., & Andrews, T. J. (2016). Patterns of neural response in scene-selective regions of the human brain are affected by low-level manipulations of spatial frequency. *Neuroimage*, *124*, 107–117.

Watson, D. M., Hartley, T., & Andrews, T. J. (*in review*). Category-selective patterns of neural response in scene-selective regions to intact and scrambled images.

Watson, D. M., Andrews, T. J., Hartley, T. (*in review*). A data driven analysis reveals the importance of image properties in the neural representation of scenes.

Results from multiple empirical chapters have been presented at the following conferences:

Watson, D. M., Andrews, T. J., & Hartley, T. (2016, January). A data-driven analysis reveals the neural representation of scenes in high-level visual cortex is linked to low-level properties of the image. *Paper presented at the meeting of the Experimental Psychology Society, London, UK.*

Watson, D. M., Hartley, T., & Andrews, T. J. (2014, September). Patterns of response in scene-selective regions of the human brain are affected by low-level manipulations of spatial frequency. *Paper presented at the meeting of the British Association for Cognitive Neuroscience, York, UK*.

Watson, D. M., Hartley, T., & Andrews, T. J. (2014, May). The topographic organization of scene-selective regions in the human brain is closely linked to the statistical properties of the image. *Poster presented at the meeting of the Vision Sciences Society, Florida, USA.*

Watson, D. M., Hartley, T., & Andrews, T. J. (2014, April). The functional organization of scene-selective cortices in the human brain is tightly linked to the statistical properties of the image. *Poster presented at the meeting of the Applied Vision Association, York, UK.*

Watson, D. M., Hartley, T., & Andrews, T. J. (2014, January). The topographic organization of scene-selective regions in the human brain is closely linked to the statistical properties of the image. *Paper presented at the meeting of the Experimental Psychology Society, London, UK.*

Hartley, T., Watson, D. M., & Andrews, T. J. (2013, November). Consistent topographic patterns of response in scene selective cortex are strongly correlated with scene-centred image statistics. *Poster presented at the meeting of the Society for Neuroscience, San Diego, California, USA.*

Chapter 1 – Literature Review

1.1 What is scene perception?

Human observers are able to perceive and extract information from visual scenes across a hugely diverse range of scene contexts and viewing conditions. In general terms, a visual scene can be considered as a view of an environment in which objects, surfaces, and textures are arranged in a manner indicating a particular spatial layout (Oliva, 2013). Importantly, this definition may encompass a highly diverse range of scene contexts, including man-made and natural scenes, and indoor and outdoor scenes. Common examples of visual scenes include room interiors, natural landscapes, and cityscapes. Our capacity to reliably perceive and extract information from visual scenes is key to our ability to successfully interact with our local spatial environment, for instance in the case of navigating a familiar route, learning new routes, and recognising landmarks.

Despite the complexity of real world visual scenes, human observers are able to reliably identify scenes even when they are presented for durations as short as mere fractions of a second (Potter, 1975; Greene & Oliva, 2009a), and when presented under severely visually degraded conditions (Torralba, 2009). How is it that the human visual system is able to extract the key components of visual scenes so efficiently? One suggestion is that scene processing may follow a coarse-to-fine (or alternatively global-to-local) processing bias in which the more global, coarse-scale features of the scene are extracted rapidly, and this information is then later complemented by a slower but more detailed analysis of the local, fine-scale components of the scene (Schyns & Oliva, 1994). Importantly, coarse-scale visual components of scenes have been noted to reliably cue the overall spatial structure of scenes – often referred to as the *gist* of the scene (Oliva & Torralba, 2001; Torralba & Oliva, 2003). Thus, key visual components of scenes may underscore human scene perception.

At the same time, recent neuroimaging studies have identified a number of regions in the human brain that appear to respond selectively to images of visual scenes (Nasr et al., 2011). These regions include the Parahippocampal Place Area (PPA; Epstein &

Kanwisher, 1998), Retrosplenial Complex (RSC; Maguire, 2001), and Occipital Place Area (OPA; Dilks et al., 2013). Together, these regions are thought to form a scene processing network within the brain, the function of which is proposed to underscore visual scene perception.

The remainder of this chapter will: 1) outline in more detail how visual statistics of scenes may be used to cue key spatial properties of scenes, 2) outline how such visual properties relate to human behaviour during scene perception, 3) overview evidence on the neural bases of scene processing and possible contributions of visual scene properties to these processes, and 4) provide an overview of the main aims and content of this thesis.

1.2 Visual statistics of scenes

One clue as to how the visual system is able to extract the spatial components of scenes so efficiently may lie in the statistical regularities present in the visual properties of scene images. Oliva & Torralba (2001) note that the low-level visual statistics of images differ both reliably and markedly across different types of scenes. Figure 1.1 depicts images taken from 5 distinct scene categories (city, coast, forest, indoor, and mountain) along with their corresponding Fourier amplitude spectra. For these purposes, the amplitude spectra can be thought of as providing a graphical method of illustrating the spectral (spatial frequency and orientation) properties of the images. Amplitude spectra have been calculated either across the whole image or within local windows of a 4x4 grid such that the spatial distribution of the spectral properties across the image can be seen. In both cases, the spectral properties can be seen to differ between the images.

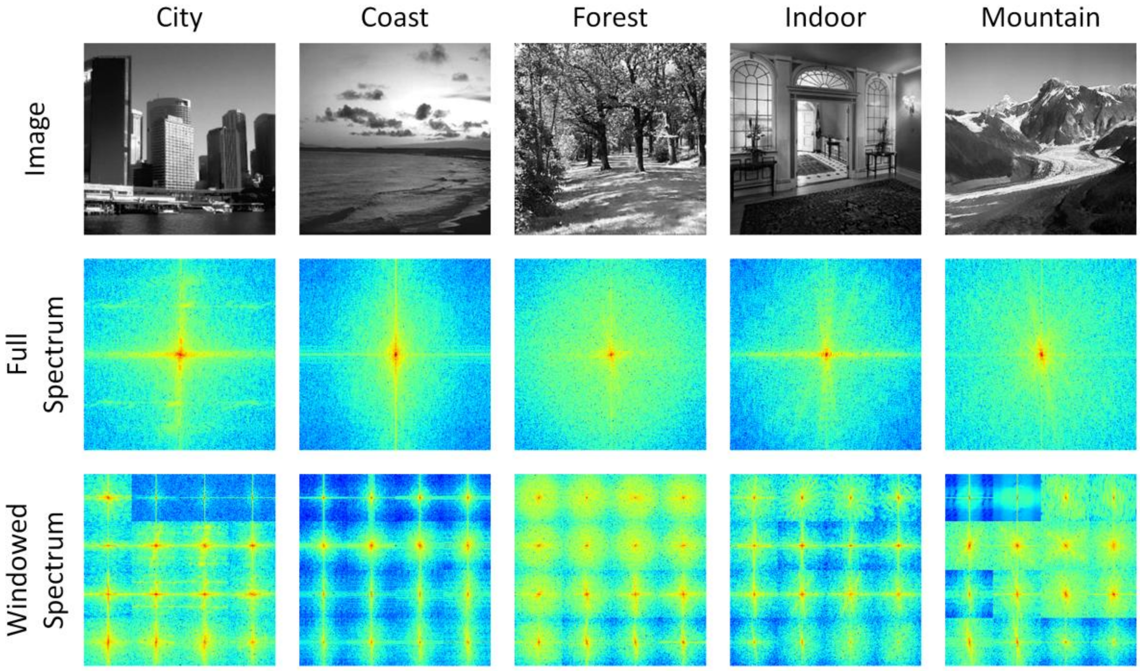


Figure 1.1 Examples of images from 5 different scenes categories (top row). The middle row shows the corresponding Fourier amplitude spectra calculated across the whole image, whilst the bottom row shows the amplitude spectra calculated within 4x4 windows of the image. In both cases, spectral statistics can be seen to differ between the scenes. In particular, the windowed spectra demonstrate how variability of the spectral properties across the spatial extent of the image predicts viewing distance. To aid viewing, amplitude spectra are displayed on a log scale.

Oliva and Torralba (2001) note that key perceptual dimensions of scenes (e.g. naturalness, openness, expansiveness, etc.) can reliably capture the spatial structure of scenes. Furthermore, they propose that these dimensions can in turn be effectively captured by visual statistics of the image, such as spectral properties (e.g. the spatial frequency and orientation content of the image) and the coarse-scale organisation of such properties across the scene. For instance, the windowed spectra in Figure 1.1 show how some scenes (e.g. forest, indoor) have relatively consistent spectral properties across the image, whilst others show more variability (e.g. city, coast, mountain). The consistency of these visual properties across a scene is often termed the stationarity (or conversely non-stationarity) of the image statistics (Torralba & Oliva, 2003). The stationarity of the statistics can predict the viewing distance of the scenes, with more distant scenes showing less stationarity due to the inclusion of more varied textures such as sky in the upper and terrain in the lower portions of the image. Oliva & Torralba (2001) suggest the notion of a *spatial envelope* of the scene – the combination of visual properties of an image that together can reliably cue the overall spatial structure or *gist* of the scene. Oliva and Torralba further demonstrate that a machine learning algorithm trained on a statistical measure of visual properties such as these can reliably discriminate images from different categories of scenes (e.g. coasts from forests). More recently, similar approaches have successfully made even finer distinctions, such as discriminating sub-categories of indoor scene (Quattoni & Torralba, 2009).

Thus, visual properties of scenes differ reliably between different types of scenes, such as different scene categories. This raises the possibility that the human visual system may exploit these statistical regularities to aid scene perception and extraction of scene *gist*.

1.3 Scene statistics relate to human behaviour

One line of evidence supporting the role of spatial envelope properties in human perception of scenes is that scene statistics have been shown to relate to behavioural measures of scene perception. Greene and Oliva (2006) tested participants' ability to categorise rapidly presented scenes. Although overall performance was high, it was observed that when participants did make miscategorisations they were often for scenes with similar spatial properties to the target category. For instance forest scenes are marked by low openness (i.e. low spatial expanse), and hence a large number of false positives when forest was the target were for non-forest scenes but which also possessed low openness. Furthermore, Greene and Oliva (2010) demonstrated behavioural adaptation to spatial envelope properties. Participants viewed a stream of rapidly presented scenes sharing similar spatial properties, and were then required to categorise a final target image with an ambiguous scene category. It was observed that participants' categorisation of the target image could be biased away from the spatial properties of the adapted images. For instance, when adapting to a stream of scenes marked by low openness followed by an ambiguous target image that could reasonably be categorised as either a forest or a field, participants would be more likely to categorise this as a field as this would be associated with higher openness.

It has also been observed that machine learning algorithms trained to categorise scenes based on their spatial envelope properties are able to reliably predict human categorisation behaviour (Greene & Oliva, 2006), and indeed outperform equivalent models based on semantic or object properties of the scenes (Greene & Oliva, 2009b). Ehinger et al. (2011) observed that such machine learning algorithms are better able to categorise scenes rated as 'typical' of their category by human observers, suggesting that scene typicality may be predicted by the degree to which a scene possesses spatial envelope properties typical of its category.

In the same way as Oliva and Torralba (2001) propose that the spatial envelope of a scene can be described by the low-level visual properties of the scene, a number of studies have investigated how specific visual dimensions may influence scene perception. For instance, studies have noted that the spatial frequency content of an image influences scene perception. Specifically, it has been suggested that scene perception follows a coarse-to-fine process (Schyns & Oliva, 1994) in which coarse scale, low spatial frequency features are extracted from the scene rapidly to give an initial first-pass analysis that then helps inform a later, more detailed processing of the finer scale, high spatial frequency components. Such a bias would permit the extraction of key spatial components of scenes faster than if all spatial frequency bands had to be processed simultaneously. Schyns & Oliva (1994) presented participants with hybrid images of scenes – composite images containing a low-pass filtered image overlaid with a separate high-pass filtered image. The image thus contains two different scenes, and importantly although the components of both scenes are always available, perception can be biased towards one or the other scene depending on which frequency band is attended to. It was found that when images were presented very rapidly (for 30ms) participants were more likely to perceive the low frequency scene, whereas with longer stimulus durations (for 150ms) participants tended to favour perception of the high frequency scene, consistent with a coarse-to-fine processing bias across time. Similarly, Kauffmann et al. (2015b) report faster reaction times during a scene categorisation task for movie sequences of a scene running from low-pass to high-pass filtered images than vice versa.

Interestingly, the coarse-to-fine processing bias may not always be absolute. For instance, Oliva & Schyns (1997) note that both the low and high spatial frequency components of a hybrid scene are able to prime recognition of a later target scene, even when presented for short stimulus durations. Thus, although the visual system may ordinarily favour a coarse-to-fine bias, both low and high frequency bands are available early on in the timecourse, and the system is flexible such that high frequency components can be extracted if required.

Meanwhile, McCotter et al. (2005) adapted the *'bubbles'* technique (Gosselin & Schyns, 2001) in order to directly investigate the spectral components most informative to scene categorisation. For a given image, a set of components of the Fourier phase spectra were sampled, and all other components replaced with random noise. After inverse transforming the spectrum back to the image domain, only the sampled components would remain intact whilst the un-sampled components would be rendered unrecognisable. Participants then performed a scene categorisation task on these images, with a different random set of components being sampled on each trial. In this way, over a large number of successive trials it was possible to measure which spectral components were more or less informative to categorising each scene category. In terms of spatial frequency, it was found that the most informative components for all categories occurred at relatively low frequencies, suggesting that coarse scale components of scenes alone provide sufficient information to accurately discriminate scene categories. This would be consistent with the coarse-to-fine processing bias (Schyns & Oliva, 1994) in that the general scene *gist* can be extracted from the coarse components rapidly, and that such coarse scale components can reliably describe the spatial envelope of a scene (Oliva & Torralba, 2001). In terms of the orientation of the visual components, McCotter et al. report that different orientations were informative for different scene categories. For instance, horizontal orientations were most informative for discriminating coastal scenes, consistent with the horizontal bias present due to the dominant horizon line within such scenes. Meanwhile, oblique orientations were more informative for discriminating mountain scenes, consistent with the sloping edges of the terrain in such scenes.

In summary, behavioural evidence suggests that the human visual system is sensitive to the spatial envelope properties of scenes during scene perception. In the same way as spatial envelope properties can be reliably described by the low-level visual properties of scenes (Oliva & Torralba, 2001), human scene perception is also influenced by the spectral components of scenes such as spatial frequency and orientation. Taken together, these results suggest that the human visual system may exploit the statistical regularities present in the visual components of scenes to aid scene perception.

1.4 Neural responses to scenes in the human brain

Recent neuroimaging studies have identified a number of regions in the human brain that appear selectively responsive to scenes or places. That is, these regions respond more to images of scenes than they do to images from other visual objects categories, such as faces or inanimate objects, or to scrambled images of scenes. These regions include the Parahippocampal Place Area (PPA; Epstein & Kanwisher, 1998), Retrosplenial Cortex (RSC; Maguire, 2001), and the Transverse Occipital Sulcus / Occipital Place Area (TOS / OPA; Dilks et al., 2013). The locations of these regions are illustrated in Figure 1.2. The selectivity of these regions for scenes appears fairly ubiquitous; for instance, they will respond preferentially across a hugely diverse range of scenes including both indoor and outdoor scenes, and man-made and natural scenes (Epstein & Kanwisher, 1998). Furthermore, patients with damage to these regions often suffer with topographagnosia, in which they exhibit severe impairments in scene recognition (especially those lacking major landmarks), novel route learning, and spatial navigation within both familiar and unfamiliar environments (Barrash et al., 2000; Maguire, 2001; Mendez & Cherrier, 2003). This suggests that these regions are closely related to our ability to perceive and interact with spatial environments.

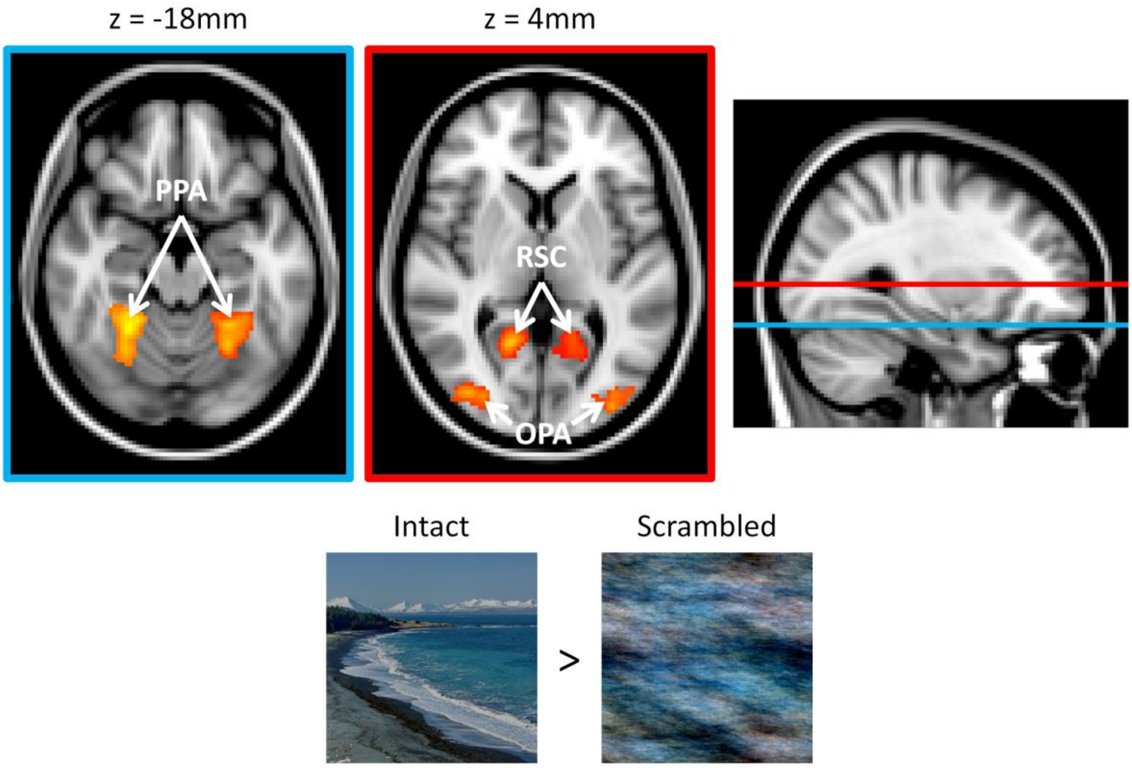


Figure 1.2 Locations of core scene selective regions, overlaid on the standard MNI152 brain. Statistical maps indicate the group average responses (N = 20) to a contrast of intact scenes over Fourier phase scrambled scenes. The Parahippocampal Place Area (PPA) is located on the ventral-temporal surface, the Retrosplenial Cortex (RSC) on the medial-temporal surface, and the Occipital Place Area (OPA; previously referred to as the Transverse Occipital Sulcus) on the lateral-occipital surface.

1.4.1 Perspectives on the neural representation of scenes

Many of the earlier studies of scene-selective cortices focussed on traditional univariate analysis methods, i.e. those that consider the amplitude of neural response. These studies identified scene selective regions as responding relatively uniformly across a wide range of scene stimuli, such as indoor and outdoor scenes, and man-made and natural scenes (Epstein & Kanwisher, 1998). However, more recent studies have begun to employ multivariate rather than univariate analyses. These methods will be overviewed in more detail in the next chapter, but in brief here - whilst univariate methods simply examine the amplitude of response on a voxel-by-voxel basis, multivariate pattern analysis methods examine the distributed pattern of response across many voxels simultaneously. In contrast to the results of the univariate methods, studies employing multivariate methods have shown patterns of response that differ reliably between different types of scenes, such as different scene categories (Walther et al., 2009, 2011). These results would suggest a more nuanced neural representation of scenes in which distributed response patterns are tied to the stimulus properties.

One key question then is which precise stimulus dimensions underlie the functional response of scene regions, i.e. which dimensions drive the regions to be scene-selective? This issue remains highly debated within the literature. Some accounts have argued for a relatively high-level organisation of scene-selective cortices in which responses are tied to the semantic features of scenes. Importantly these representations are suggested to be largely dissociated from the visual features of the image. For instance, Walther et al., (2009) note distinct patterns of neural response in scene-selective regions to different semantic categories of scene (e.g. beaches, buildings, mountains, etc.), suggesting an organisation based upon categorical principles. This leads them to conclude that "[the] representation of scenes in higher visual areas, namely PPA, RSC, and LOC, more closely tracks human behaviour rather than physical similarity". See also Walther et al. (2011) and Stansbury et al. (2013) for similar accounts.

However, other studies have proposed a more mid-level representation that stresses a role for the spatial envelope properties of scenes. Using an event-related fMRI design, Kravitz et al. (2011) presented participants with a range of scene images. Importantly, by measuring responses on an image-by-image basis, this design avoided the constraint of grouping images by scene category at the stimulus presentation stage. Kravitz et al. found that response patterns in scene selective regions differed reliably between scenes as a function of the spatial expanse of the images (i.e. the degree to which the scene appeared open or closed), but found little evidence for responses grouping by semantic category. Similarly, Park et al. (2011) note a greater effect of spatial expanse than semantic content on neural responses, whilst Park et al. (2015) note effects of both spatial expanse and visual clutter. Previously reported effects of scene category (Walther et al., 2009, 2011) may therefore have reflected responses to spatial envelope properties that also differ reliably between scene categories, rather than a direct encoding of the semantic category *per se*.

Meanwhile, other accounts have proposed even lower-level representations that are tied to the visual features of scenes. For instance, many studies have reported biases for multiple low-level visual properties including a retinotopic bias (Malach et al., 2002; Arcaro et al., 2009; Silson et al., 2015), and biases for spatial frequency (Rajimehr et al., 2011; Kauffmann et al., 2014), visual orientation (Nasr & Tootell, 2012), and rectilinearity (Nasr et al., 2014). Importantly, such accounts do not dispute that these regions are scene-selective, but rather suggest that scene-selectivity may arise from the interaction of multiple low-level biases that are themselves predictive of scenes.

A complication in this debate is that many of the aforementioned features across low-, mid-, and high-level accounts are themselves correlated, making it difficult to disentangle the effects of any one account from the others (Lescroart et al., 2015). For instance, it has already been discussed how low-level, visual features can be used to reliably predict the mid-level, spatial envelope properties of scenes, and how these in turn can predict the high-level, semantic category of scenes (Oliva & Torralba, 2001; Torralba & Oliva, 2003). Thus the stimulus dimensions underlying the functional organisation of scene selective cortices remain controversial, and require further investigation. One of the key aims of this thesis is to further test the role of higher- versus lower-level stimulus features in the neural representation of scenes.

1.4.2 Scene-selective regions

Whilst a number of scene-selective regions in the human brain have been identified, it has also been suggested that each region may play subtly different roles in scene perception (Epstein, 2008). The following sections review the literature on each of the scene-selective regions in more detail.

1.4.2.1 Parahippocampal Place Area (PPA)

In an early functional magnetic resonance imaging (fMRI) study, Aguirre & D'Esposito (1997) reported a region in ventral-temporal cortex that seemed to respond preferentially during a scene categorisation task. Epstein & Kanwisher (1998) more definitively showed selectivity of a bilateral region in the posterior parahippocampal gyri to images of visual scenes. In light of its location, they named this region the Parahippocampal Place Area (PPA). More recent reports have localised the PPA as spanning the parahippocampal gyrus and collateral sulcus (Nasr et al., 2011).

It has been suggested that the PPA is primarily concerned with encoding the local spatial geometries of scenes (Epstein, 2008). For instance, Epstein & Kanwisher (1998) note that the PPA response was reduced when images were fragmented and re-arranged (thereby disrupting the spatial layout of the scenes) suggesting a scene-centred representation that focuses on spatial layout. Epstein et al. (1999) report the PPA responds preferentially even to Lego models of scenes. Model scenes will emulate the spatial geometries present in real scenes, but will clearly lack the wider scene context provided by real scenes. Henderson et al. (2011) report greater PPA responses to scenes that convey a strong sense of 3D spatial structure (e.g. rooms) than those that do not (e.g. cityscapes). As previously discussed, studies employing multivariate analyses have also suggested that the PPA encodes the spatial expanse of scenes (Kravitz et al., 2011; Park et al., 2011, 2015).

It should, however, be noted that an alternative hypothesis has been proposed by Bar and colleagues (Bar et al., 2008a, 2008b) suggesting that the PPA processes contextual associations which simply coincide with scene processing, rather than being specifically scene selective *per se*. In support of this, Bar et al. (2008a) report stronger PPA responses for familiar than unfamiliar faces, despite the fact that these clearly do not embody any spatial components. Meanwhile, Bar et al. (2008b) report greater PPA activity to objects with strong contextual associations to scenes (e.g. traffic lights) than those with weak associations (e.g. a water bottle), suggesting a more object-centred representation. However, Epstein & Ward (2010) provide a rebuttal to this hypothesis. They note that Bar et al. (2008b) used a relatively slow presentation rate that might have allowed for visual imagery of scenes, especially in the strong association condition. Epstein & Ward show that with faster presentation rates that reduce the time available for any mental imagery to occur, the effect of context drops out. Furthermore, they were entirely unable to replicate the findings of Bar et al. (2008a). Thus, on the whole the literature primarily supports a scene-centred role of the PPA in encoding the spatial geometries of scenes.

A number of studies have identified other more specific aspects of PPA response properties. Firstly, there exists conflicting evidence on the effects of scene familiarity, with Epstein et al. (1999) reporting no significant effect of familiarity, whilst Epstein et al. (2007b) report greater responses to familiar than unfamiliar scenes. Epstein et al. (2007b) suggest that their earlier report may have suffered from having too small a sample size.

In terms of effects of viewpoint, the evidence is again conflicting. Epstein et al. (2003) used an fMRI adaptation paradigm in which participants viewed blocks of scenes where each block showed the same image repeatedly either from the same or different viewpoints. It was found that whilst the PPA adapted to scenes shown from the same viewpoint, it displayed a release from adaptation when scenes were shown from different viewpoints. In fact, the response to different viewpoints of the same scene was almost identical to the response to entirely different scenes. This would suggest that the PPA is sensitive to the viewpoint of the scene, and essentially fails to discriminate the same scene viewed from a different viewpoint from an entirely different scene. Conversely however, Ewbank et al. (2005) in fact do report adaption to scenes even when viewed from different angles. They suggest their discrepant findings may be due to their using smaller viewpoint shifts than Epstein et al. (2003). A further possible explanation may lie in later reports that suggest viewpoint effects may interact with the familiarity of the scene, with an increasing degree of viewpoint invariance observed as familiarity with a given scene increases (Epstein et al., 2005, 2007b).

Meanwhile, other studies have argued for lower-level representations of scenes more closely tied to the visual features of the image. For instance, a number of studies have suggested a retinotopic bias in PPA. A series of studies by Malach and colleagues have suggested a peripheral bias in scene-selective regions including the PPA (Levy et al., 2001, 2004; Hasson et al., 2002, 2003; Malach et al., 2002). Malach *et al.* suggest this bias may aid scene processing as real world scenes often span a large extent of the visual field and hence extend greatly into the periphery. This contrasts with regions selective for faces and words which they report showing a foveal bias, consistent with our tendency to fixate such objects. Furthermore, Arcaro and colleagues have identified two retinotopic maps in posterior parahippocampal cortex named as PHC-1 and PHC-2 (Arcaro et al., 2009; Wang et al., 2015) which heavily overlap with functional definitions of the PPA. Each of these maps contains a representation of the contralateral visual field and, consistent with Malach *et al.*, show an overrepresentation of the periphery. Silson et al. (2015) used a population receptive field (pRF) mapping technique in conjunction with fMRI to demonstrate the presence of both a contralateral and an upper visual field bias in PPA. Similarly, MacEvoy & Epstein (2007) note a contralateral visual field bias in that greater responses are seen in the PPA contralateral to the field of unilaterally presented stimuli. Interestingly, MacEvoy & Epstein do note that despite this contralateral bias, some bilateral response is still seen, and that fMRI adaptation can be seen bilaterally at approximately equal magnitudes. This suggests that receptive fields in the PPA may be relatively large and therefore cross the visual midline. This is consistent with more recent pRF mapping studies that have reported relatively large receptive fields both in the PPA and in ventral-temporal cortex in general (approximately 3 degrees of visual angle) compared to early visual regions (Kay et al., 2015; Silson et al., 2015).

Further supporting a lower-level representation of scenes in the PPA, a number of other studies have identified response biases to several other visual features of scenes. For instance, Rajimehr et al. (2011) report a bias towards high over low spatial frequency content in both scene and non-scene stimuli. They suggest this may coincide with a higher degree of high spatial frequencies in scene stimuli relative to other stimuli such as faces, or may help enhance edge detection processes that could be relevant for extracting spatial geometries. Furthermore, Musel et al. (2014) note greater PPA responses to movie sequences depicting a scene moving from a coarse (low-pass filtered) to a fine (high-pass filtered) scale than a fine-to-coarse scale, consistent with reports of a coarse-to-fine bias in the behavioural literature (Schyns & Oliva, 1994; Oliva & Schyns, 1997; Kauffmann et al., 2015b). It is not entirely clear how an overarching bias for high-spatial frequencies would support a further coarse-to-fine processing bias, although it should be noted that Rajimehr et al.'s (2011) study measured neural responses using fMRI, which possesses relatively poor temporal resolution, whilst the coarse-to-fine processing bias is thought to occur on a much faster temporal scale. Further studies employing more temporally sensitive methods such as EEG/MEG may be required to resolve this issue. Meanwhile, Nasr & Tootell (2012) report a bias towards cardinal over oblique orientations, again for both scene and non-scene stimuli. More recently, Nasr et al. (2014) further suggest the presence of a rectilinear bias, i.e. for rectangular visual features with straight edges as opposed to more rounded features. It is suggested that such biases reflect the relatively high occurrence of such features in natural scenes.

1.4.2.2 Retrosplenial Cortex (RSC)

The retrosplenial cortex (RSC) is located superior to the PPA on the medial temporal surface, anterior to the calcarine sulcus and posterior to the corpus callosum (Maguire, 2001; Vann et al., 2009; Nasr et al., 2011). The term “retrosplenial complex” is sometimes alternatively used to acknowledge that functional definitions of the area may encompass a number of anatomical regions. Much like the PPA, the RSC is also reported to respond preferentially to visual scenes, but is also implicated in other aspects of scene processing such as spatial memory and navigation (Maguire, 2001; Vann et al., 2009).

A number of the response properties of the RSC overlap with those of the PPA. Much like the PPA, Epstein et al. (2007b) note effects of scene familiarity on RSC responses, and increasing viewpoint invariance with increasing familiarity. Henderson et al. (2011) again note stronger RSC responses to scenes that convey a greater sense of local 3D depth, much the same as the PPA. Meanwhile, Nasr et al. (2014) also find evidence for a rectilinear bias in RSC, although they do note that this bias is more evident for visually complex stimuli, whilst the effect in the PPA was more ubiquitous.

Nevertheless, there are also a number of functional distinctions between the RSC and PPA. The main difference is that the RSC appears much more heavily implicated in aspects of spatial memory and navigation than the PPA. Epstein & Higgins (2007) note greater RSC responses to images of scenes which are accompanied with a label denoting the scene context (e.g. “beach”) than those without labels, whilst the PPA failed to show a significant difference between these conditions. A number of reports have identified greater RSC responses when participants are required to identify the location of a scene, compared to tasks requiring scene category or viewpoint discriminations (Epstein & Higgins, 2007; Epstein et al., 2007a). Furthermore, the RSC appears highly responsive during tasks involving spatial navigation (Maguire, 2001), whilst Schinazi & Epstein (2010) implicate the region in real-world route learning. It has been suggested that the RSC may represent the environment in terms of local spatial reference frames, for example encoding local heading direction and position within the immediate environment (Vass & Epstein, 2013; Marchette et al., 2014).

On the basis of such results, Epstein and colleagues have proposed distinct but complimentary roles for the RSC and PPA (Epstein & Higgins, 2007; Epstein et al., 2007a; Epstein, 2008). Specifically, it is proposed that whilst the PPA is primarily concerned with extracting the spatial geometries of the immediate local spatial environment, the RSC then focuses on attempting to locate this scene within the wider spatial environment. A further, but similar hypothesis is that the RSC may act as a mediator that transforms between egocentric representations of the environment (i.e. those centred with regards to the observer) in visual and parietal regions, and allocentric representations of the environment (i.e. those centred with regards to the wider environment) in medial temporal and hippocampal regions (Burgess et al., 2001; Vann et al., 2009).

1.4.2.3 Transverse Occipital Sulcus (TOS) / Occipital Place Area (OPA)

The transverse occipital sulcus (TOS) is located on the lateral occipital surface, overlapping retinotopic regions V3B, V7, and LO1 (Grill-Spector, 2003; Nasr et al., 2011). Recently, Dilks et al. (2013) have suggested renaming the region as the occipital place area (OPA) in order to better reflect its functional rather than anatomical definition, and the fact that more recent estimates have actually placed its location slightly outside of the anatomically defined TOS (Nasr et al., 2011). Consequently, the term OPA is now becoming more prevalent within the literature.

In comparison to the PPA and RSC, much less is known about the response properties of the OPA. Much like the PPA and RSC, the OPA responds preferentially to images of scenes (Nasr et al., 2011). Studies applying transcranial magnetic stimulation (TMS) to the OPA have reported disruptions to behavioural performance on both scene categorisation and scene matching tasks, but little to no effect on task performance with images of objects or faces, thus demonstrating a causal role of the OPA in scene perception (Dilks et al., 2013; Ganaden et al., 2013). Similar to the PPA and RSC, Epstein et al. (2007b) report scene familiarity effects within the OPA, in addition to an increasing degree of viewpoint independence with increasing scene familiarity. Nasr et al. (2014) also find a rectilinear bias within this region.

A number of studies have suggested retinotopic biases within this region, consistent with its location overlapping known retinotopic regions (Nasr et al., 2011). The peripheral bias reported in the PPA is also observed in the OPA (Malach et al., 2002). Meanwhile, Silson et al.'s (2015) pRF-mapping study reported both contralateral and lower visual field biases. Silson et al. also report relatively large receptive field sizes of approximately 3 degrees of visual angle, comparable to those in the PPA.

The functional role of the OPA within the scene processing network is poorly understood. Dilks et al. (2013) suggest it may represent an initial stage within a hierarchical scene processing network, analogous to proposed roles of the occipital face area in the face processing network (Haxby et al., 2002). However, this conclusion is largely based on its posterior location and proximity to the occipital face area, and currently is lacking much direct empirical evidence. Meanwhile Silson et al. (2015) propose an alternative hypothesis that the PPA and OPA may represent complimentary regions that function in parallel with one another. This is based on their observation that the PPA and OPA display upper and lower visual field biases respectively, and they suggest that this may represent a natural continuation of the upper and lower visual field segregation between ventral and lateral pathways observed in early visual areas, for example between V2v/V2d and V3v/V3d (Wandell & Winawer, 2011). Nevertheless, this is the only study to have investigated this hypothesis, and so again further empirical evidence is required.

1.4.2.4 Other scene-selective regions

Although the PPA, RSC, and OPA are typically regarded as forming the core components of a scene processing network, it should be noted that other brain regions are also implicated in scene processing. Whilst the PPA is located in the posterior portions of the parahippocampal gyrus, scene selectivity can be seen to extend further anterior along parahippocampal cortex into entorhinal cortex and the hippocampus. Anterior parahippocampal cortex has been implicated in wayfinding and novel route learning in scenes (Janzen & Weststeijn, 2007; Janzen et al., 2007; Janzen & Jansen, 2010; Wegman & Janzen, 2011). Furthermore, Epstein (2008) notes that whilst lesions to posterior parahippocampal cortex may impair navigation and route learning in both new and familiar environments, lesions to anterior parahippocampal cortex are more likely to be associated with impairments only in new environments. Moving further anterior still, both entorhinal cortex and the hippocampus have been implicated in processes relating to navigation and spatial memory, such as encoding an organism's current heading direction and spatial location within the environment (Hartley et al., 2014). Thus, moving from posterior to anterior there seems to be a general progression in scene selectivity from more visually based processes in PPA / posterior parahippocampal cortex towards more navigationally relevant and memory based processing in anterior parahippocampal cortex, entorhinal cortex, and the hippocampus.

1.4.3 Connectivity of scene-selective regions

With regards to the PPA, Kim et al. (2006) used Diffusion Tensor Imaging (DTI) to measure the structural connectivity of the PPA with other regions in both early and higher visual cortices. They reported a high density of white matter tracts between the PPA and early visual regions – in particular V1, V2, V3v, and V4 – but very few connections to other high level visual regions (FFA, LO, and hMT+). Interestingly, the other high level visual regions did show a high degree of connectivity to one another, as well as the early visual regions. This suggests that in terms of feed-forward input from early visual cortices, the PPA may exist on a separate neural pathway distinct from those which connect other high level visual regions. However, Mullin & Steeves (2013) report that applying TMS to object-selective lateral-occipital cortex (region LO) disrupts PPA responses to scenes as measured by an immediately subsequent fMRI scan. This would suggest that some functional connectivity between PPA and LO exists, even if direct anatomical connections are not present. Other studies have noted functional connectivity between the PPA and frontal / parietal regions (Kauffmann et al., 2015a, 2015c). It has been suggested that a PPA – fronto-parietal network may support a coarse-to-fine process of scene recognition (Kauffmann et al., 2014). Specifically, it is proposed that low spatial frequency, coarse scale components receive a rapid first-pass analysis, fed forward to fronto-parietal regions in order to quickly extract the overall spatial components of the scene. This information can then be fed back to visual regions such as the PPA in order to better inform a later parsing of the high spatial frequency, fine scale features. Finally, Baldassano et al. (2013) suggest a possible anterior-posterior division in PPA connectivity. Using measures of functional connectivity, they show connectivity between posterior PPA and lateral-occipital visual areas such as OPA and LO, whilst anterior regions showed greater connectivity with RSC and parietal regions. They suggest that posterior PPA may be more concerned with representing visual properties of the stimulus, whilst anterior regions might be more associated with higher level processes such as spatial memory.

In terms of RSC connectivity, Kobayashi & Amaral (2007) note a very high density of projections from RSC to both hippocampal and parahippocampal regions in the macaque brain. Although less dense, they also note some projection to frontal cortices. Consistent with these results, a DTI study in humans reported a large number of white matter tracts between RSC and regions of the medial temporal lobe, including the hippocampus (Greicius et al., 2009). Meanwhile, Kim et al. (2015) report the case of a patient with congenital topographagnosia who displayed reduced functional connectivity between the PPA and RSC relative to healthy controls. Taken together, these results suggest the RSC is heavily connected with hippocampal and parahippocampal regions. These results are therefore consistent with the proposed functional roles of the RSC both in terms of complimenting PPA function (Epstein, 2008) and in mediating between hippocampal and other brain regions (Burgess et al., 2001).

In contrast to the PPA and RSC, relatively little information is available on the connectivity of the OPA with other regions, and what literature does exist often appears conflicting. When applying TMS to the OPA, Mullin & Steeves (2013) did not observe any significant effects on the fMRI activation within the PPA, suggesting against direct connectivity with the PPA. However, in their study of a congenital topographagnosic, Kim et al. (2015) did report reduced functional connectivity between OPA and PPA, although the magnitude of this effect was less than in the connectivity between PPA and RSC. Baldassano et al. (2013) reported significant functional connectivity between the OPA and PPA, although they do note this relationship is more prevalent for posterior than anterior PPA. Thus, the connectivity of the OPA is an area requiring further investigation.

1.4.4 Feedback and lateral inputs to scene-selective regions

Much of the evidence discussed thus far has considered the feed forward modulation of responses in scene-selective regions. Nevertheless, the extensive feedback as well as feedforward connections present throughout the visual system make it seem likely that scene selective regions can be modulated by top-down influences just as any other visual region can. Indeed, task demands have been shown to modulate the neural response to visual objects in both ventral-temporal and early visual cortices (Harel et al., 2014). Furthermore, it also remains possible that scene-selective regions may be modulated by lateral input from other cortical regions. For instance, Wolbers et al. (2011) report PPA responses to haptic input from touching Lego models of scenes in both sighted and congenitally blind participants, raising the exciting possibility of cross-modal input to scene selective cortices. Thus it remains possible, if not likely, that whilst scene selective cortices may be strongly driven by bottom-up visual input, responses can nevertheless also be modulated by other stimulus properties via top-down and lateral inputs.

1.4.5 A unifying model of category selectivity

So far we have seen that there exist regions in the human brain that respond selectively to scenes, and that such regions may be functionally organised along a number of biases for various stimulus properties. However, many of these biases are relatively weak, and it remains unclear how such strong category selectivity could arise from these. For instance, whilst scene selective regions may display a particular bias for the spatial frequency content of an image, the magnitude of this bias is considerably less than that of the categorical response of these regions to scenes over other visual object categories. Op de Beeck et al. (2008) have proposed one particularly influential model that aims to explain how strong, localised selectivity may arise from weak biases for underlying stimulus properties. This model is illustrated in Figure 1.3. Their proposal is that there are a number of topographically organised maps that each encodes a particular functional dimension of the stimulus in a distributed manner across the cortical surface. For instance, one map could encode a spatial frequency bias, another an orientation bias, another a retinotopic bias, and so on. Each map on its own may exhibit only weak selectivity, but crucially the spatial organisation of each map is correlated with that of the other maps. For instance, the points in the spatial frequency map that are most predictive of scene images would overlap with the points in the orientation and retinotopy maps that are also predictive of scenes. If responses are combined across maps, for instance by multiplication, then each of the underlying weakly biased maps can give rise to a strong, localised peak of activity. Importantly, this model does not dispute the notion that cortical regions may exhibit category selectivity, e.g. for scenes. Rather, it simply proposes a mechanism by which such selectivity could occur, i.e. that it arises from a combination of distributed and topographically organised biases for a number of stimulus properties that are themselves predictive of that stimulus class.

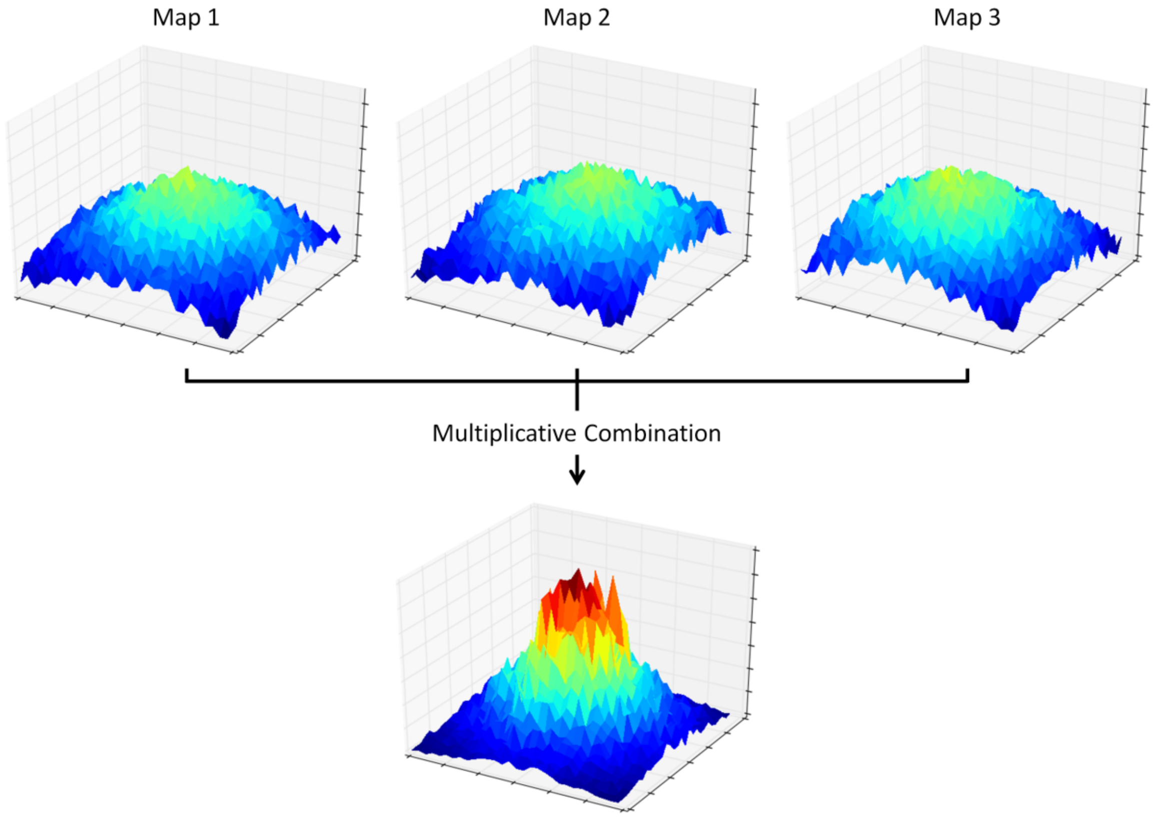


Figure 1.3 Illustration of Op de Beeck et al.'s (2008) model proposing how localised functional selectivity may arise from widely distributed responses. A series of functional maps exist distributed across the cortical surface. Each map exhibits a bias for a particular stimulus property that is relevant to encoding that stimulus class. Each map on its own may exhibit only a weak bias. However, the organisation of each map is spatially correlated with the other maps; for instance, the peak in selectivity in one map for a particular stimulus class spatially coincides with the peak in selectivity in the other maps for that same stimulus class. Combination of the maps (e.g. by multiplication) therefore produces a strong and localised peak in selectivity for that stimulus class.

1.5 The GIST descriptor: measuring scene statistics

Given the apparent importance of spatial envelope properties to both human scene perception and the neural representation of scenes, it is important to be able to empirically measure such scene statistics. This thesis focuses on one particularly influential model of scene statistics: the GIST descriptor (Oliva & Torralba, 2001; Torralba & Oliva, 2003), so named as it attempts to capture the key spatial properties of scenes that are extracted during the rapid perception of scene *gist*. The GIST descriptor represents an image in terms of its spectral properties (spatial frequency and orientation) and the spatial distribution of these across the extent of the image. This process is illustrated for an example image in Figure 1.4. GIST descriptors can be calculated for many images, and then used to make statistical comparisons between images. Importantly, and in validation of the GIST descriptor, the visual properties that are measured are the same types of features that have been proposed to effectively describe spatial envelope properties (Oliva & Torralba, 2001). Furthermore, these visual properties have been related to both human scene perception (Schyns & Oliva, 1994; McCotter et al., 2005), and the neural representation of scenes (Arcaro et al., 2009; Rajimehr et al., 2011; Nasr & Tootell, 2012). Indeed, the GIST descriptor has been shown to reliably predict neural responses to visual objects in ventral temporal cortex (Rice et al., 2014; Andrews et al., 2015). The GIST descriptor has also proved practically applicable within the field of computer vision; for instance, see Pugeault & Bowden (2011) for a successful application to an algorithm for a self-driving car.

It should be noted that the GIST descriptor is by no means the only statistical image descriptor available. A non-exhaustive list of other popular image descriptors includes the HMAX model (Riesenhuber & Poggio, 1999), SIFT descriptor (Lowe, 2004), and HOG descriptor (Dalal & Triggs, 2005), whilst other more recent approaches have shown success with deep neural-network learning models (Szegedy et al., 2015). Meanwhile, other studies have reported advantages of combining multiple image descriptors (e.g. Xiao et al., 2010). However, it is beyond the scope of this thesis to provide a full comparison of image descriptors. Instead, this thesis focuses on the application of the GIST descriptor to predicting neural responses to scenes. It does therefore remain possible that an alternative model may prove a better predictor of responses than the GIST descriptor. However, this is not necessarily problematic as it is not the goal of this thesis to identify the best image descriptor for predicting responses. Rather the aim is simply to test whether the GIST, as an example of a neurologically plausible model of scene statistics, is able to predict neural responses. If it can, then this would suggest that scene selective regions of the human brain are sensitive to the spatial envelope properties that the GIST captures.

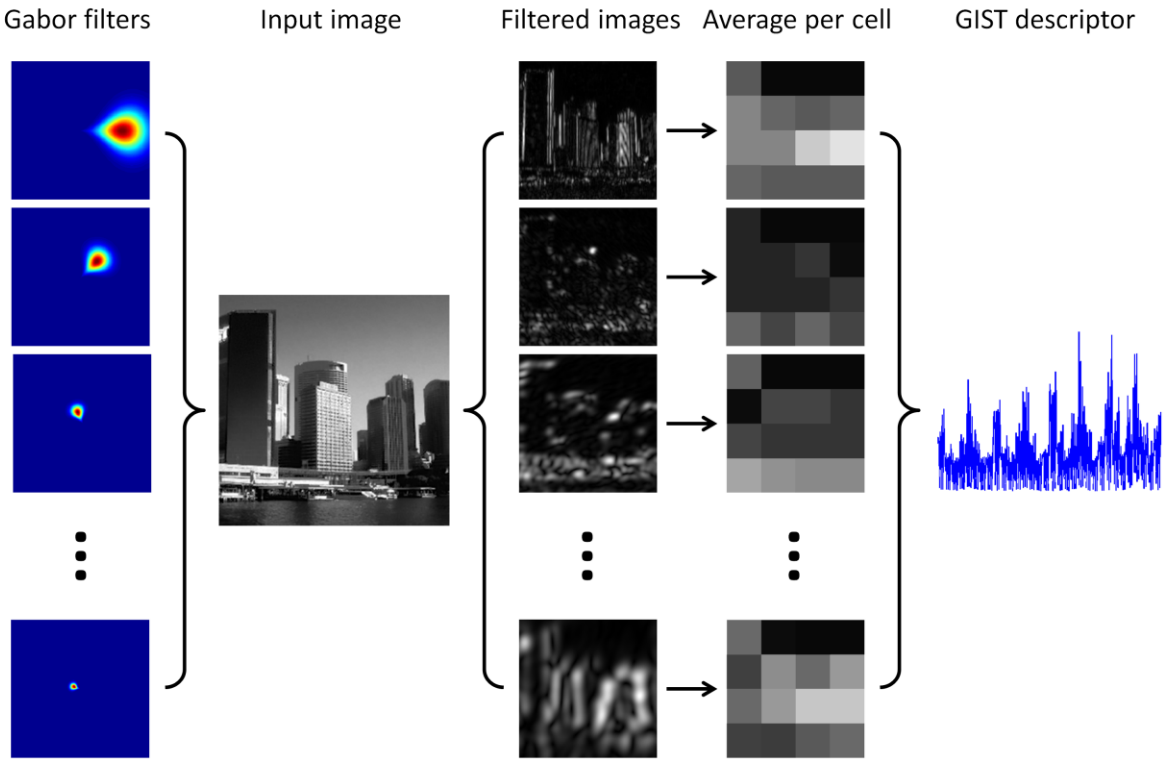


Figure 1.4. Illustration of the calculation of a GIST descriptor (Oliva & Torralba, 2001) for an example image. The image is first convolved with a bank of 32 Gabor filters spanning 4 spatial frequencies and 8 orientations. In order to capture the spatial distribution of the spectral properties across the image, each of the resulting filtered images is then downsampled to a 4x4 grid and the pixel intensities averaged within each grid cell. The GIST descriptor is constructed by reshaping each of these to a 16x1 vector and concatenating them across the 32 filters, yielding a final 512x1 vector that describes the image in terms of the spatial frequencies and orientations present at each of the 16 locations across the image.

1.6 Thesis overview

This literature review has demonstrated how key spatial properties of scenes can be described by the low-level visual statistics of images represented by the spatial envelope of the scene. These spatial envelope properties have been shown to predict both human behaviour in scene perception and the neural representations of scenes. Finally, the GIST descriptor provides a statistical method of capturing the spatial envelope properties of scenes.

Nevertheless, there do remain a number of outstanding issues. It remains unclear to what extent neural responses in scene selective regions are governed by higher level categorical principles in comparison to lower level visual dimensions. Furthermore, whilst a number of studies have identified biases for low level visual features in scene selective regions, these studies have typically employed univariate analyses that simply measure the amplitude of response. Thus, it remains unclear to what extent these features may be represented in the distributed patterns of response that are assessed by multivariate methods. Using fMRI in conjunction with multivariate pattern analyses, this thesis therefore aims to address the following questions:

1. Can the low-level visual properties of scenes predict patterns of neural response to such images in scene selective cortices?
2. What are the relative contributions of low-level visual information versus high-level semantic category information to the neural response?
3. What are the contributions of specific low-level visual properties (spatial frequency, orientation, and retinotopy) to the neural response?
4. Are there alternatives to the more traditional categorical accounts that might more parsimoniously explain the function of scene selective cortices?

The second chapter provides a more detailed overview of the fMRI methods applied within this thesis, and in particular those relating to multivariate pattern analyses.

The third chapter describes two fMRI experiments that test whether the visual statistics of scenes as measured by the GIST descriptor can predict the neural response patterns to different categories of scene.

The fourth chapter describes two further fMRI experiments that provide a direct manipulation of two key visual properties of scenes (spatial frequency and orientation). The effects of these visual properties on the neural response patterns are directly compared and contrasted to those of scene category.

The fifth chapter describes an fMRI experiment that examines how the retinotopic distribution of visual features across the scene affects the neural response under conditions where perception of the scene category is impaired. This is accomplished by measuring patterns of neural response to images from different scene categories which have been either globally or locally scrambled to disrupt scene perception.

The sixth chapter describes a final fMRI experiment that aims to address concerns about the assumptions of categorical scene structure made by many previous studies. Here an entirely data-driven method is used to objectively cluster scenes by their visual properties, as measured by the GIST descriptor. Multivariate pattern analyses in conjunction with fMRI are then used to test whether neural responses to these visual scene clusters can be discriminated.

The seventh chapter overviews the key findings of the thesis, and discusses how they relate to theories of the neural representation of scenes in the human brain.

Chapter 2 – fMRI Methods Review

2.1 The fMRI BOLD signal

Functional Magnetic Resonance Imaging (fMRI) is a commonly used method for measuring neural responses in the brain *in vivo*. As neurons increase their firing rate, they consume their energy reserves which then need replenishing. The brain accomplishes this by increasing the transfer of oxygen to those cells via the bloodstream, causing a local change in the oxygenation of the blood. Ogawa et al. (1990) proposed using MRI to provide a measure of this oxygenation change using a contrast termed the Blood Oxygenation Level-Dependent (BOLD). In this way, vascular changes in the brain as measured by fMRI via the BOLD signal can be used to infer the underlying neural activity. The BOLD signal is therefore the fundamental measurement of almost all fMRI research. A number of statistical techniques have been proposed for analysing such data; here we focus on two of the most commonly used groups of methods: univariate and multivariate analyses.

2.2 Univariate Analysis

Traditional fMRI analyses have tended to employ a univariate general linear model (GLM) approach (Friston et al., 1995) to analysing the BOLD signal. A set of regressors are defined that model the neural response to different stimulus conditions. For instance a boxcar model, which predicts zero response when the stimulus is absent and a non-zero response when the stimulus is present, can be convolved with a hemodynamic response function to produce an expected timeseries of response. This model can then be regressed against the genuine fMRI BOLD signal on a voxel-by-voxel basis. This results in a whole-brain statistical map of parameter estimates (regression coefficients) that reflect the fit of the model; a voxel that is responsive to the stimulus will be predicted well by the model and hence will be assigned a large coefficient value, whilst one that is not responsive will be predicted poorly and assigned a smaller coefficient value. This process is illustrated in Figure 2.1. This analysis can be performed for a number of experimental conditions, and if desired these conditions can be contrasted against one another. From here, the statistical significance at each voxel can be estimated (e.g. as a p-value, or a z-score) by testing the coefficient values against a baseline. Such statistical maps can then be submitted to further statistical analysis as necessary, such as higher-level analyses for combining results across scan sessions and / or subjects.

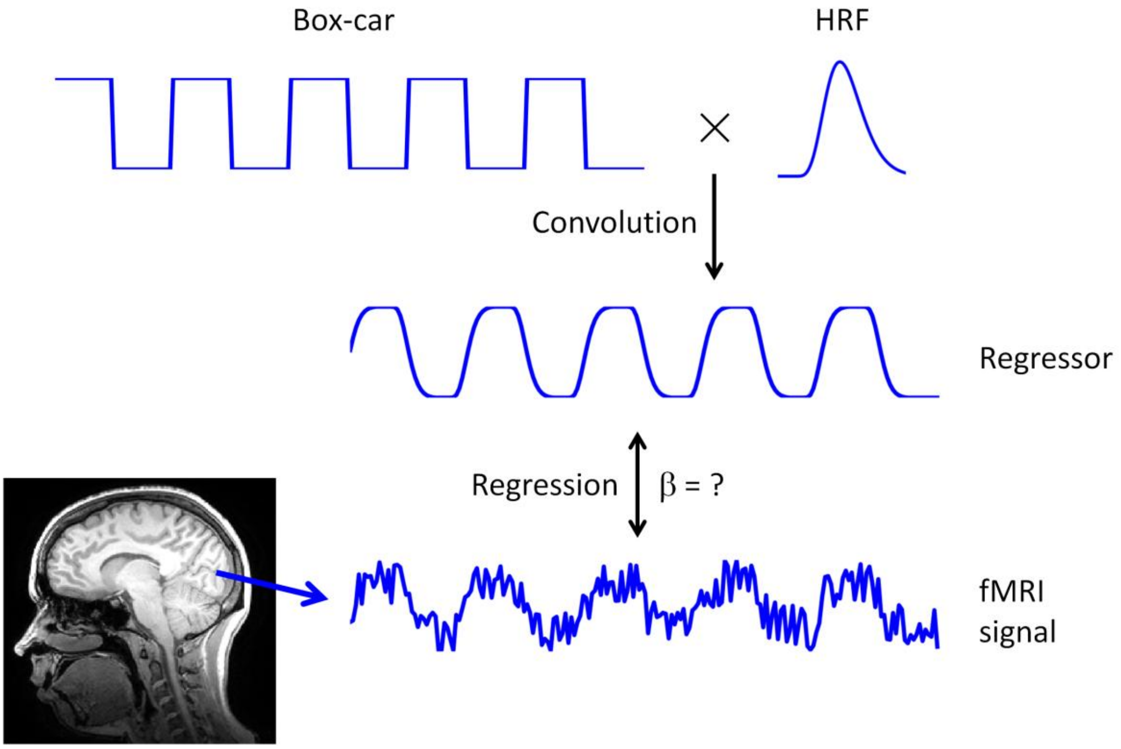


Figure 2.1. Example of univariate GLM analysis of fMRI data. A box-car function is defined that corresponds to the periods of stimulus presentation. A hemodynamic-response function (HRF; in this instance, a single-gamma function) is then convolved with the box-car to produce a hemodynamic regressor. This can then be regressed against the fMRI signal for each voxel independently. This results in a statistical map of parameter estimates that indicate the fit of the regressor to the fMRI BOLD signal at each voxel.

Although the univariate GLM approach is undoubtedly a powerful statistical technique for the analysis of fMRI data, it does nevertheless have limitations. The technique simply estimates the amplitude of neural response on a voxel-by-voxel basis, and it is assumed that responses which deviate significantly from zero reflect stimulus related neural activity. However, it is not necessarily the case that a sub-threshold or near-zero response means that a voxel does not convey information about the stimulus. For instance, coherent patterns of neural response may be observed across multiple voxels which may include voxels showing both super- and sub-threshold positive and negative responses. Neural patterns such as these may be found to differ reliably between stimulus conditions. In such cases it may even be that the aggregate response across voxels in a given brain region is near zero, but crucially this does not mean that this region does not contain information about the stimulus. Standard univariate analyses, however, will not be sensitive to information represented in distributed neural patterns. It is for this reason that there is now a growing use of multivariate methods within neuroimaging research that aim to capture precisely this type of information.

2.3 Multivariate Pattern Analysis (MVPA)

An alternative to the standard univariate approach is the use of multivariate methods that allow one to consider the patterns of response across multiple voxels simultaneously. Such approaches are often grouped under the term of Multi-Voxel or Multi-Variate Pattern Analysis (MVPA). MVPA methods therefore provide a different sensitivity to traditional univariate analyses. The input to a MVPA is simply any measure of the patterns of neural response. It is possible to perform the MVPA on the fMRI timeseries, however it is often more common to use the outputs of an initial univariate analysis, such as the parameter estimate statistical maps. It should be noted that although MVPA techniques are discussed in relation to fMRI here, these methods are equally applicable to other neuroimaging methods such as EEG or MEG (e.g. see Carlson et al., 2013; Cichy et al., 2014).

2.3.1 Correlation-based methods

One of the simplest forms of MVPA is the correlation method, which is the method that was applied in the original Haxby et al. (2001) study that first proposed the application of MVPA to fMRI data. It is common practice to perform MVPA across a subset of voxels rather than the whole-brain volume, for instance by restricting the analyses to a pre-defined region of interest (ROI). Firstly, estimates of the patterns of response are generated for each stimulus condition. Typically, these are the parameter estimate values generated by a univariate GLM analysis for each voxel in a region. To determine if these patterns of response are reliable, it is necessary to cross-validate the analysis, i.e. compare conditions across independent estimates of the neural responses. To this end, multiple estimates of each condition are generated. For instance, parameter estimates could be generated separately for odd and even runs of the stimulus presentation for each condition. Although these patterns could be entered into the MVPA directly, at this point they are likely to still contain a high degree of variance that is shared across the conditions and may not necessarily reflect the key stimulus dimensions of interest, such as generic responses to visual stimulation or attentional effects. In order to better isolate the unique variance to each condition, it is common practice to perform some normalisation of the data. A typical normalisation procedure is to subtract from each stimulus condition a per-voxel mean across all the stimulus conditions – the logic being that the mean should reflect the shared variance across conditions, so subtracting this should reduce the influence of this shared variance whilst leaving the unique variance relatively unaffected. In some cases, this normalisation is taken a step further by also dividing by a per-voxel estimate of the standard deviation across conditions such that responses are converted to z-scores. In order to ensure that splits of the cross-validation remain independent, normalisation should be performed within each split independently (Kriegeskorte et al., 2009).

Once the normalised patterns of response have been obtained, one can now proceed with the main MVPA. Pairwise correlations are calculated between the neural response patterns for each possible combination of conditions across the splits of the data. This is performed for both within-condition comparisons (e.g. faces-even with faces-odd) and between-condition comparisons (e.g. faces-even with houses-odd). This process is illustrated in Figure 2.2a. If the total number of comparisons is large, it is often convenient to represent the results within a correlations matrix (Figure 2.2b). The prediction is that if response patterns can discriminate the stimulus conditions, then the within-condition correlations should be higher than the corresponding between-condition correlations. This result would indicate that patterns of response to a given stimulus condition are more similar to other responses from that same condition than to responses from other conditions, and from this we can infer that there is some information represented in these distributed neural response patterns that allows us to discriminate the conditions.

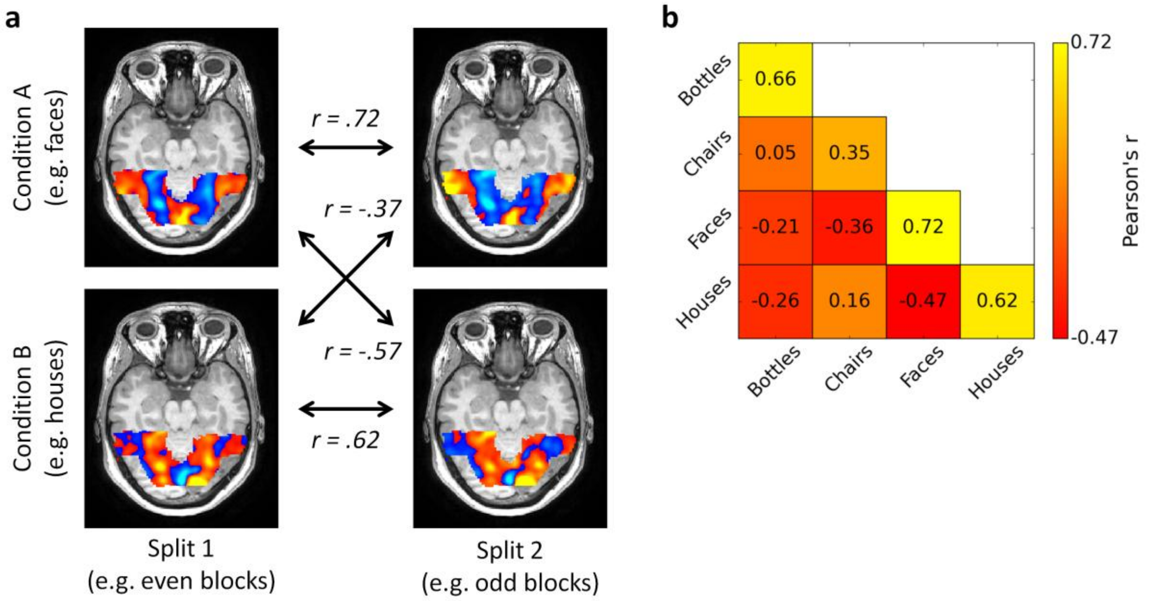


Figure 2.2. Illustration of correlation-based MVPA paradigm. (a) Patterns are estimated for two stimulus conditions (e.g. faces and houses), each across two independent splits of the data (e.g. even and odd stimulus runs). Patterns are restricted to a region of interest (ROI), and correlated pairwise both within- and between-conditions across the data splits. Higher within- than between-condition correlations indicate patterns can be discriminated. (b) For a larger number of comparisons, results may be more easily represented in a correlations matrix; within-condition comparisons are represented on the diagonal elements, between-condition comparisons are represented on the off-diagonal elements. In this example, off-diagonal elements symmetrically opposite one another across the diagonal (e.g. faces-even/houses-odd and faces-odd/houses-even) have been averaged to aid visualisation.

In Haxby et al.'s (2001) study, participants were presented with images from 8 visual object categories: faces, houses, cats, bottles, scissors, shoes, chairs, and scrambled images. A univariate GLM approach was used to generate parameter estimates for each of the stimulus conditions relative to rest, with patterns estimated for odd and even runs of the stimulus presentation independently. Parameter estimates were normalised by subtracting a voxel-wise mean across all conditions within each split independently. Analyses were restricted to an ROI of ventro-temporal cortex, and responses correlated across the splits. Haxby et al. demonstrated higher within- than between-category correlations for all stimulus conditions, indicating that response patterns could be discriminated. Furthermore, it was found that this result held even when the analysis was restricted to only those voxels which showed the maximum univariate response to each of the stimulus conditions, and also in the reverse case where these voxels were excluded from the main analysis. This demonstrates that reliable information about all the stimulus conditions was present in the neural response patterns both within and outside of the regions maximally responsive to each of the stimulus conditions. Importantly, it is unlikely that a standard univariate analysis would have been sensitive to this information. For instance, simply aggregating response amplitudes across the face selective voxels would likely have revealed strong responses to faces, but then just uniformly weak responses to each of the other stimulus categories. By contrast, the multivariate analysis was able to reveal the information present within the distributed neural response patterns in these regions.

2.3.2 Classification algorithms

Shortly after Haxby et al.'s (2001) original proposal of correlation-based MVPA, alternative strategies were proposed that employed classification algorithms derived from the machine-learning literature. These algorithms are more sophisticated than the correlation method and have the potential to provide greater sensitivity. Even Haxby et al.’s data received a re-analysis using a neural-network algorithm (Hanson et al., 2004). The initial stages of analysis are similar to those of the correlation method. Once again, it is necessary to cross-validate the analysis in order to test how well the model will generalise to new examples. To this end, estimates of the patterns of response are generated for multiple independent splits of the data. It is also common to normalise the data such as by mean subtraction or z-scoring, as per the correlation method. However, the implementation of the pattern analysis itself differs substantially from the correlation methods.

Many different classification algorithms are available, with popular examples including the k-nearest-neighbour classifier, linear discriminant analysis, and the support vector machine (Mur et al., 2009). However, all of these algorithms work along a similar principle. Each sample in the dataset is represented as a point within a (potentially high-dimensional) feature space. The features of this space most commonly correspond to voxels in the brain, but can equally well correspond to other dimensions; for instance O'Toole et al. (2005) use a feature space based on principal components derived from the voxel activity. Generally speaking the number of samples available in fMRI data is relatively small compared to the total number of voxels in the brain. It is not advisable to perform classification in a feature space with many more features than samples, thus it is necessary to perform *feature selection* in order to reduce the number of features initially selected. Common approaches to feature selection include restricting analyses to a pre-defined ROI, selecting a subsample of the most strongly modulated voxels, using a searchlight approach (see below), reducing dimensionality with principal components analysis, or any combination of these (Mur et al., 2009).

The general form of the classification paradigm is illustrated in Figure 2.3. Here, we imagine we have a very small ROI comprising just two voxels; hence samples are represented in a 2D feature space. Samples each belong to one of two classes, for instance faces and houses. The first step is to train the classification algorithm on a subset of the data given by the cross-validation paradigm (for instance, a leave-one-run-out cross-validation would use data from all but one of the stimulus runs). The classification algorithm will attempt to find a decision boundary that optimally separates samples between the classes; the precise definition of the "optimal" boundary depends on the choice of classifier. Typically decision boundaries are linear, although some algorithms (e.g. some support vector machines) offer the option to use non-linear decision boundaries instead. Non-linear decision boundaries allow the possibility to model more complex stimulus relationships. However, the potential benefits of this can also be negated by the added sensitivity leading the algorithm to be more susceptible to overfitting, making it poorer at generalising to new examples in the testing-phase of the cross-validation (Misaki et al., 2010).

Regardless of the choice of classifier or type of decision boundary, once the classifier has been trained it can then be used to make predictions about the class membership of samples. If a given sample falls on one side of the decision boundary we predict that it is a member of one class, and if it falls on the other side we predict that it belongs to the other class. Testing classification accuracy on the training set is likely to lead to inflated estimates of the accuracy due to overfitting effects. Instead, we will typically test performance on an independent subset of the data not previously seen by the classifier (for instance, a leave-one-run-out cross-validation would use data from the excluded run). In this way, we can test the ability of the classifier to generalise to new examples, which is more useful than simply knowing how well it fits its own specific training data. If the class we predict for a given sample matches its actual class then this counts as a correct classification. If we predict a different class than the true class then this is a misclassification. This entire procedure can then be repeated for the remaining folds of the cross-validation (for instance using each stimulus run as the excluded run once in a leave-one-run-out cross-validation). If we are able to successfully discriminate the classes from one another based on the neural patterns across our features, we would expect overall above chance classification accuracy.

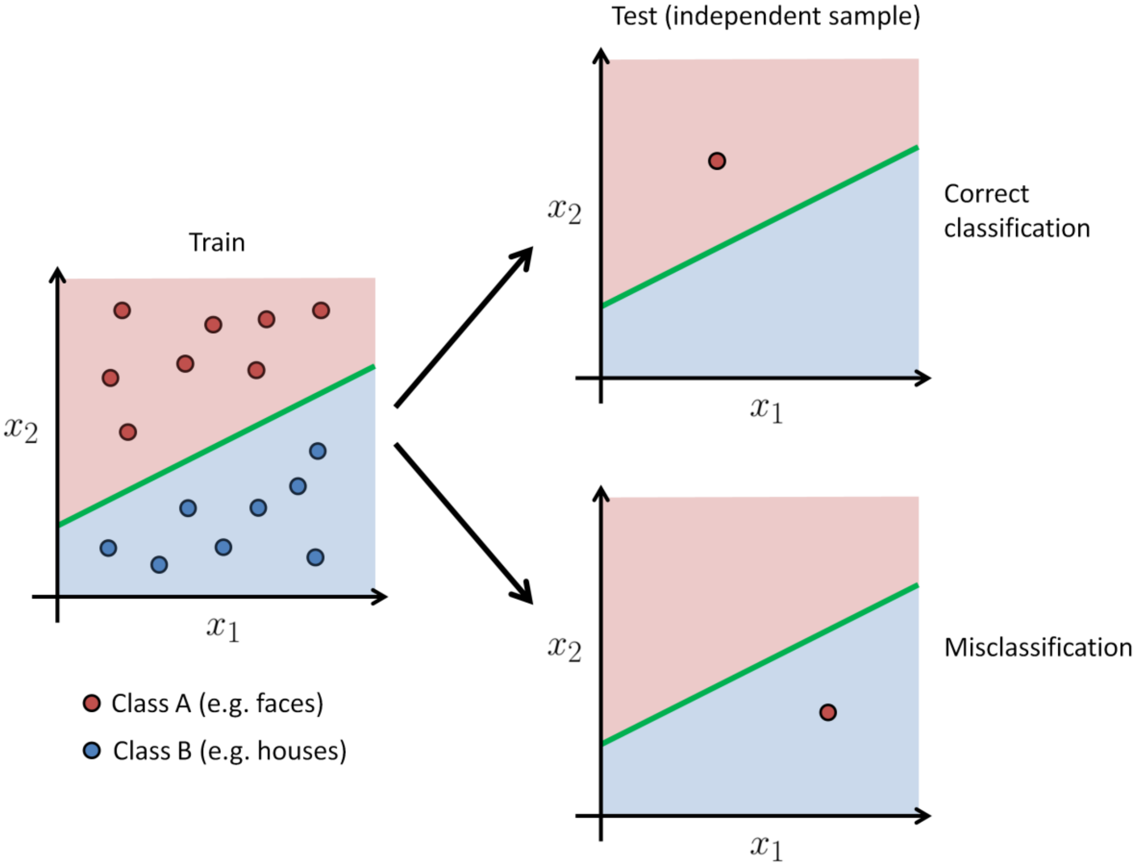


Figure 2.3. Illustration of classification-based MVPA paradigm. In this example, responses to two classes of stimuli (e.g. faces and houses) are measured across two voxels (labelled as features x_1 and x_2). Each response is represented as a sample within the feature space. A classification algorithm is trained on a subset of the data (e.g. on all but one of the stimulus runs) to place a decision boundary (green line) that optimally separates the two classes. In this example, any samples falling above the decision boundary (red shaded region) will be classified as Class A, whilst those falling below the decision boundary (blue shaded region) will be classified as Class B. Classification accuracy is assessed by testing the classifier on an independent subset of the data not included in the training set (e.g. the left out stimulus run). This process may then be repeated for the remaining folds of the cross-validation scheme.

In this example, we considered the case where we had just two voxels / features. In reality, we would often have many more features than this, in which case our samples become represented in a hyper-dimensional space. However, the underlying principle of the classification algorithm remains the same. We also only considered the case where there are two classes. It is also possible to perform classification with multiple classes, in which case most implementations will attempt to break the task down into a series of two-class classification problems, such as by using a one-versus-all (in which classifiers are trained to discriminate each class in turn from the concatenation of all other classes) or a one-versus-one approach (in which classifiers are trained to discriminate each pairwise combination of classes in turn).

Although classification algorithms offer the possibility of greater sensitivity than correlation methods (Mur et al., 2009), they do also themselves present a number of methodological problems. As previously mentioned, classification algorithms may prove unreliable if the number of features greatly exceeds the number of samples. Unfortunately, due to the poor temporal resolution of fMRI, data acquisition tends to be relatively slow and as a result the number of obtainable samples is often quite small. Feature selection can help reduce the number of features, but the number is still likely to exceed the number of samples. Consequently, classification algorithms often require many orders of magnitude more data to be collected than for the equivalent correlation analyses simply to obtain a sufficient number of samples. Whereas a correlation analysis could potentially be run on a single fMRI-scan run, a classification algorithm is likely to require many repeated scan runs. A further issue relates to the information that is recoverable from the algorithm. Imagine a hypothetical case where a classification algorithm is trained to discriminate 3 classes – A, B, and C. Responses to class A are found to be most similar to other responses from class A, but are still fairly similar to responses from class B, and not at all similar to responses from class C. Nevertheless, classes A and B are still sufficiently dissimilar that a classifier is able to successfully discriminate them. The classifier only reports what classifications are made, so the information that A was still relatively similar to B is lost. This information will only be evident if A and B are sufficiently similar that they become confusable by the classifier, i.e. all responses that are confusable must be similar, but not vice versa. Thus, although classification algorithms may offer greater sensitivity at discriminating responses than correlation methods, they are also potentially less sensitive at modelling the relationships between conditions. If the primary goal of the analysis is to determine if conditions can be discriminated, and if the data are appropriate for classification (e.g. a sufficient number of samples can be acquired), then classification algorithms may be the preferable option. On the other hand if the goal is to also model the relationship between conditions, for instance for a representational similarity analysis (see below), or if the data are not suitable for classification, then a correlation method may be more appropriate.

2.3.3 Representational (Dis)similarity Analysis (RSA / RDA)

Whilst both correlation and classification methods can be used to determine if response patterns can be discriminated, a limitation of such analyses is that it often remains unclear what functional dimensions underlie these responses. An alternative application of MVPA that has grown in popularity in recent years is representational similarity / dissimilarity analysis (RSA / RDA; Kriegeskorte et al., 2008; Nili et al., 2014), which aims to provide an explicit test of the functional dimensions underlying neural responses.

The general form of the RSA paradigm is illustrated in Figure 2.4. Two or more similarity matrices are constructed, for instance by taking the pairwise correlations between each combination of conditions. Alternatively dissimilarity matrices can be constructed, for instance by calculating one minus the correlation, in which case a representational dissimilarity analysis (RDA) is conducted. However, this choice is largely arbitrary as the final outcome of the analysis will be the same between RSA and RDA. RDA may be more appropriate if one wishes to submit the matrices to further analyses that require distance rather than similarity measures, such as hierarchical clustering or multi-dimensional scaling (Nili et al., 2014). Similarity / dissimilarity matrices can be constructed from any data available provided they all correspond to the same set of conditions; in the example illustrated, one is constructed from neural data via correlation-based MVPA (see above), and another from a model of the stimulus. It is also possible to use the confusion matrices derived from classification algorithms (e.g. see O'Toole et al., 2005; Walther et al., 2009), however a more continuous measure such as correlation is usually preferable as it is likely to be more sensitive to the relationships between conditions.

Once constructed, the similarity / dissimilarity matrices are then compared to one another in turn, for instance by correlating the elements between matrices. If a high degree of similarity is seen between two matrices this indicates that each measure is able to predict the relative similarity between the conditions in the other measure. From this we can infer that similar functional dimensions underlie both measures. In this way, we can explicitly test hypotheses about the dimensions underlying the neural response.

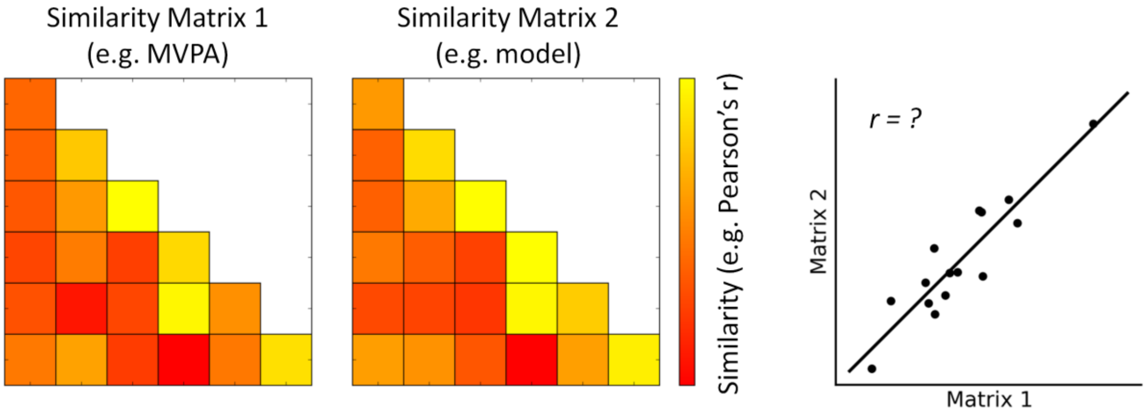


Figure 2.4. Illustration of the Representational Similarity Analysis (RSA) paradigm (Kriegeskorte et al. 2008). Two similarity matrices are constructed, for example by calculating pairwise correlations between each combination of conditions. Similarity matrices can be constructed from any data available provided they both correspond to the same stimulus conditions; in this example, one is determined from neural data via MVPA, and the other from a model of the stimuli. Representational similarity is assessed by comparing the similarity of the two matrices, for instance by correlating them. High similarity between the matrices indicates that each measure is a good predictor of the relative similarity between conditions in the other measure.

2.3.4 Searchlights

The multivariate methods discussed thus far have all required some form of pre-defined feature selection, such as by restricting analyses to a pre-selected ROI. However, this necessarily constrains the information that we can derive about the spatial location of the multivariate information within the brain. For instance, it may remain unclear whether all voxels within a given region carry critical information, or only a few. Furthermore, we cannot determine if significant information is present in other brain regions outside of our voxel selection. An alternative implementation of MVPA that attempts to address these issues is the searchlight approach proposed by Kriegeskorte et al. (2006). In this analysis a small, spherical ROI is defined, and the desired MVPA procedure performed within this ROI as per normal. This procedure should return a value, for instance a within-condition minus between-condition difference in the case of a correlation-based approach, a decoding accuracy from a classification algorithm, or a correlation value from a RSA / RDA. This value is assigned into the central voxel of the sphere, and then the entire process is repeated, iterating the sphere around the whole-brain volume till every voxel has been used as the central voxel once. This results in a whole-brain statistical map where the value at each voxel reflects the result of the MVPA in a spherical ROI centred on that voxel. In this way, we can see where in the brain the multivariate information is present without needing to restrict analyses to a pre-defined set of voxels. It is also possible to perform surface-based searchlights in which circular discs are defined along the cortical surface rather than spheres in the volume, which may present some advantages over the standard volumetric method (Oosterhof et al., 2011).

However, searchlights do themselves present a number of problems. The relatively large number of spheres required to cover a whole-brain volume makes the process highly computationally expensive, although a fast implementation using a Gaussian Naive Bayes classifier has been proposed (Pereira & Botvinick, 2011) and may present some further advantages over other algorithms (Raizada & Lee, 2013). A further issue arises in combining searchlight maps across individuals. Many studies employ simple parametric tests to compare values across subjects at each voxel against chance level. However, these values may not be normally distributed so it is unclear whether these tests are appropriate. An alternative is to determine significance via permutation testing, however the already considerable computational cost of the searchlight multiplied by the time required to run a large number of permutations at each sphere makes this approach largely unfeasible. A possible solution is to perform a much smaller number of permutations at the individual level, but then estimate a null distribution by repeatedly bootstrapping these permutations across subjects (Stelzer et al., 2013). The searchlight technique also produces a large problem of multiple comparisons due to the many thousands of statistical analyses computed across the whole brain. Although several techniques exist for performing correction for multiple comparisons in univariate data (e.g. voxel-wise or cluster-based thresholding), these techniques may not be appropriate for use with searchlight data as they often make statistical assumptions about the data being analysed. An alternative proposal is to perform cluster-based thresholding via permutation testing (Stelzer et al., 2013).

Further issues arise from the interpretation of searchlight results. It is frequently tempting to interpret the searchlight result at a given voxel as representing the information present within that voxel, in the same way as one would interpret a univariate statistical map. However, this is incorrect; in reality the value represents the information present within a sphere centred on that voxel. This distinction is subtle, but it can lead to a number of issues in the interpretation (or misinterpretation) of searchlight results (Etzel et al., 2013). Searchlight results are inherently linked to the size of the sphere used. By definition, response patterns that occur over a coarser spatial scale than the extent of the sphere cannot be considered by the searchlight. Larger spheres may allow for consideration of coarser scale patterns, and the greater number of voxels included in the sphere sample may also improve the signal-to-noise ratio. However, they may also lead to poorer spatial specificity as a cluster of informative voxels may drive high performance in any sphere overlapping it, even those that are not centred on the cluster.

In summary, searchlights offer a powerful statistical technique for applying MVPA to the whole-brain. However, they also present a number of methodological problems that need to be considered, and the results of searchlights must be interpreted with care.

2.3.5 Properties of neural patterns

The pattern information used by multivariate analyses is often thought to occur at a fine spatial scale and to be largely idiosyncratic within individuals (Haxby et al., 2014). Consequently, it is common practice to perform MVPA on fMRI data without first applying spatial smoothing in order to preserve the fine spatial scale – in contrast to traditional univariate analyses where spatial smoothing is common practice. Furthermore, analyses are typically performed within individual subjects independently, and the results then aggregated across subjects, to account for idiosyncrasies within the patterns. However, there are counter-examples to these assumptions within the literature.

In support of the notion that multivariate methods are sensitive to information occurring at a fine spatial scale, studies have shown that visual orientation can be decoded from early visual regions such as V1 (Haynes & Rees, 2005; Kamitani & Tong, 2005). These results seem counterintuitive as the orientation columns in early visual cortex occur at a very fine spatial scale, far beyond the resolution of the voxels used in fMRI. The biased sampling account proposes that by chance each voxel in turn may contain more or fewer columns sensitive to some sets of orientations than others, leading to small biases in the orientation responses of each voxel which are detectable by a pattern classifier (Haynes, 2015). This would therefore suggest that pattern information occurs at a very fine voxel-by-voxel scale. However, this conclusion seems to run counter to the relatively coarse spatial sampling provided by fMRI. Indeed, even if no spatial smoothing is applied by the experimenter as part of the data analysis, some smoothing is nevertheless inherent in the data due to the low spatial specificity of the hemodynamic response (at least relative to the scale of neuronal orientation columns) and head motion artifacts. Meanwhile, the decoding of orientation has alternatively been explained in terms of a much coarser scale bias for radial orientations (i.e. those pointing towards the fovea) across V1 (Freeman et al., 2011). Furthermore, if pattern information does occur at a fine spatial scale, one would expect spatial smoothing to significantly disrupt the information available. However, Op de Beeck (2010) demonstrates that both when decoding orientation information from V1 and when decoding object category information from lateral-occipital visual cortex, spatial smoothing not only fails to produce detrimental effects to decoding performance but in some cases actually benefits it. Thus, although the information used by multivariate analyses is often assumed to occur at a fine-scale, there is also counter evidence suggesting that at least some information may occur at coarser scales. Although not commonly practiced, applying spatial smoothing to fMRI data prior to multivariate analyses may not be as detrimental to performance as is often assumed.

A further frequently stated assumption of pattern information is that it is largely idiosyncratic to each subject (Haxby et al., 2014), and consequently analyses are typically performed within each subject independently. In particular, if pattern information does occur at a fine spatial scale (although see discussion above) then this may be expected to show poor spatial alignment across subjects. Although Haxby et al. (2001) suggest that the spatial topography of the patterns may be similar across individuals, they do not perform their correlation analyses across subjects as they suggest current inter-subject co-registration techniques are not sufficient for aligning data at a sufficiently fine scale.

Nevertheless, other studies have successfully performed cross-subject pattern analyses across a diverse range of contexts including decoding object categories (Shinkareva et al., 2008, 2011; Haxby et al., 2011), cognitive states (Mourão-Miranda et al., 2005; Poldrack et al., 2009), truth telling (Davatzikos et al., 2005), social cues (Clithero et al., 2011), and somatosensory information (Kaplan & Meyer, 2012). Studies comparing within-subject to cross-subject classification have typically reported either comparable (Kaplan & Meyer, 2012) performance, or an advantage for within-subject analyses (Davatzikos et al., 2005; Shinkareva et al., 2008, 2011; Clithero et al., 2011; Haxby et al., 2011). However, performance of cross-subject analyses in previous studies might be impeded by the fact that spatial smoothing was not applied to the individual subject data; unsmoothed patterns may be expected to show poorer alignment across subjects than smoothed patterns. Indeed, Mourão-Miranda et al. (2005) report improved performance of cross-subject analyses with spatial smoothing. Alternatively, if standard anatomical alignment of data across subjects still proves insufficient, Haxby et al. (2011) propose a "hyper-alignment" method that aligns subjects based on their functional neural responses rather than anatomy, and which may achieve better performance than standard anatomical alignment.

2.4 Overview of Thesis Methods

This thesis presents a number of fMRI experiments using MVPA to test the contribution of visual properties to the neural representation of visual scenes. In all cases, the inputs to the pattern analyses are parameter estimate maps for each condition generated by GLM univariate analyses. All experiments make frequent use of representational similarity analyses (Kriegeskorte et al., 2008; Nili et al., 2014), and consequently correlation-based analyses are employed over classification algorithms as these provide a better measure of the relationships between conditions. Chapter 3 provides a comparison of the effects of using smoothed and unsmoothed data, and of performing within-subject and cross-subject analyses. Consistent with Op de Beeck (2010), the data demonstrate a beneficial rather than detrimental effect of spatial smoothing. Furthermore, the cross-subject analyses are frequently comparable to or outperform the equivalent within-subject analyses, especially in the case where spatial smoothing is first applied to the data. Consequently, subsequent experimental chapters make exclusive use of spatially smoothed data and cross-subject analyses.

Chapter 3 – Patterns of Response to Visual Scenes are Linked to the Low-Level Properties of the Image

This chapter is adapted from: Watson, D. M., Hartley, T., & Andrews, T. J. (2014). Patterns of response to visual scenes are linked to the low-level properties of the image. *NeuroImage*, *99*, 402–410. [1]

3.1 Abstract

Scene-selective regions in the brain play an important role in the way that we navigate through our visual environment. However, the principles that govern the organization of these regions are not fully understood. For example, it is not clear whether patterns of response in scene-selective regions are linked to high-level semantic category or to low-level spatial structure in scenes. To address this issue, we used multivariate pattern analysis with fMRI to compare patterns of response to different categories of scenes. Although we found distinct patterns of neural response to each category of scene, the magnitude of the within-category similarity varied across different scenes. To determine whether this variation in the categorical response to scenes could reflect variation in the low-level image properties, we measured the similarity of images from each category of scene. Although we found that the low-level properties of images from each category were more similar to each other than to other categories of scenes, we also found that the magnitude of the within-category similarity varied across different scenes. Finally, we compared variation in the neural response to different categories of scenes with corresponding variation in the low-level image properties. We found a strong positive correlation between the similarity in the patterns of neural response to different scenes and the similarity in the image properties. Together, these results suggest that

[1] The author, David Watson, designed the experiment, analysed the results, and wrote the article under the supervision of Dr. Tom Hartley and Prof. Timothy Andrews.

categorical patterns of response to scenes are linked to the low-level properties of the images.

3.2 Introduction

The ability to perceive and recognize different visual scenes is essential for spatial navigation in the world. Although real-world scenes can be incredibly complex and heterogeneous, human observers are able to reliably recognize and categorize images of scenes even when the images are shown briefly (Potter, 1975; Joubert et al., 2007; Greene & Oliva, 2009a). These studies have been taken to suggest that the initial perception of natural images is based on the global, visual properties - the *gist* - of the scene (Oliva & Torralba, 2001; Greene & Oliva, 2009a).

Neuroimaging studies have found a number of regions of the human brain that respond selectively to visual scenes. Damage to these regions often leads to impairments that are specific to scene perception and spatial navigation (Aguirre & D'Esposito, 1999; Mendez & Cherrier, 2003). The parahippocampal place area (PPA) is a region of the posterior parahippocampal gyrus that displays preferential activity to images of scenes over and above images of objects and faces (Aguirre & D'Esposito, 1997; Epstein & Kanwisher, 1998). Other place selective regions include the Retrosplenial Complex (RSC) located immediately superior to the PPA and the Transverse Occipital Sulcus (TOS) or Occipital Place Area (OPA) on the lateral surface of the occipital lobe (Epstein, 2008; Dilks et al., 2013).

The spatial layout of different categories of scenes can vary quite considerably (Torralba & Oliva, 2003). Although neuroimaging studies using univariate analyses have reported comparable levels of response to scenes as diverse as natural landscapes, cityscapes and indoor scenes in scene-selective regions (Aguirre & D'Esposito, 1997; Epstein & Kanwisher, 1998), more recent studies using multivariate analyses have found distinct patterns of response in these regions to different categories of scene (Walther et al., 2009, 2011). Interestingly, these patterns of neural response have also been shown to correlate with patterns of behavioural response, but *not* with the low-level image properties of the images (Walther et al., 2009). This suggests that there is a dissociation between the perceptual categorization of scenes and their underlying image properties. However, this conclusion has been challenged by other studies that have suggested that the patterns of response in scene-selective regions are better explained by the spatial layout of the scene rather than by semantic category (Kravitz et al., 2011; Park et al., 2011). Although these studies are not explicit about how the image properties of the scene are linked to the patterns of neural response, work in computer vision indicates that semantically-distinct scene categories can be identified on the basis of their characteristic low-level image statistics. For example, the GIST descriptor can be used to accurately classify different scene categories and derive spatial properties such as openness (Torralba & Oliva, 2003).

Our aim was to determine whether categorical patterns of brain activity within scene-selective regions are linked to the low-level properties of the images from each category of scene. To address this issue, we measured the pattern of response to different categories of scenes using fMRI. Next, we asked how similar the low-level properties of images from each category were to each other. Finally, we asked whether differences in the categorical response to different visual scenes might be due to variation in low-level image properties. Our prediction was that, if low-level visual properties are linked to categorical patterns of response in these regions, then scene categories with similar image statistics should elicit correspondingly similar patterns of brain activity.

3.3 Methods

3.3.1 Participants

20 participants took part in Experiment 1 (9 males, mean age: 24.5) and 20 participants took part in Experiment 2 (9 males, mean age: 25.2). All participants were neurologically healthy, right-handed, and had normal or corrected-to-normal vision. Written consent was obtained for all participants and the study was approved by the York Neuroimaging Centre Ethics Committee.

3.3.2 Stimuli

All images were taken from the LabelMe scene database (http://cvcl.mit.edu/database.htm; Oliva & Torralba, 2001) and presented in greyscale at a resolution of 256 x 256 pixels. All further image processing was performed in MATLAB v7.10 (http://www.mathworks.co.uk/). Fourier-scrambled images were created by randomising the phase of each 2-dimensional frequency in the original image while keeping the power of the components constant. For each experiment, the luminance histogram of images across all conditions was equated using the SHINE toolbox (Willenbockel et al., 2010).

3.3.3 Experimental Design

In Experiment 1 and Experiment 2, participants viewed images from 5 stimulus conditions. Figure 3.1 shows examples of images taken from the stimulus conditions used in both experiments. The stimulus conditions in Experiment 1 included: 1) cityscapes, 2) indoor scenes, 3) natural landscapes, 4) mixed (interleaved images from conditions 1-3) and 5) scrambled (Fourier scrambled versions of the mixed condition). The stimulus conditions in Experiment 2 included: 1) coast, 2) forest, 3) mountains, 4) mixed (interleaved images from conditions 1-3) and 5) scrambled (Fourier scrambled versions of the mixed condition). In each experiment, images from each condition were presented in a block design with 9 images in each block. Each image was presented for 850ms followed by a 150ms black screen. Each stimulus block was separated by a 9s period in which a fixation cross was superimposed on a grey screen that was equal in mean luminance to the scene images. Each condition was repeated 8 times in a counterbalanced block design, giving a total of 40 blocks. To maintain attention throughout the scan session, participants performed a one-back task in which one image from each block was repeated. Stimuli were presented using PsychoPy (Peirce, 2007, 2009).

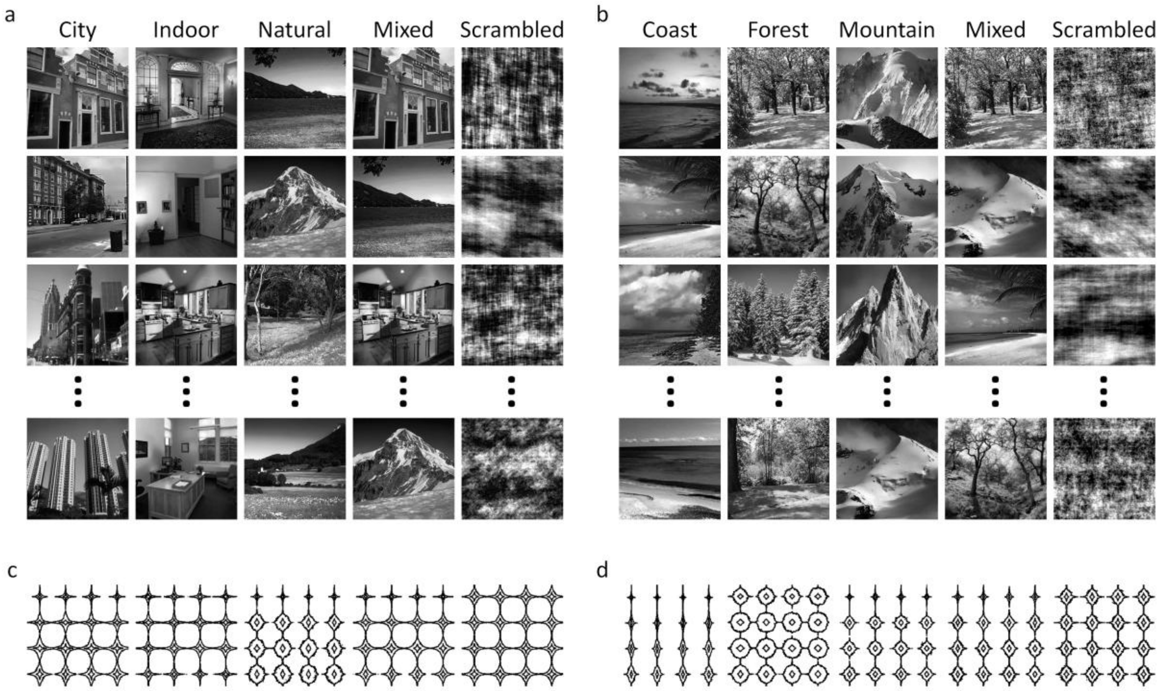


Figure 3.1. Examples of images from each experimental condition in (a) Experiment 1 and (b) Experiment 2. Category average contour plots of Fourier power spectra within 4x4 windows are shown for (c) Experiment 1 and (d) Experiment 2.

3.3.4 Imaging Parameters

All scanning was conducted at the York Neuroimaging Centre (YNiC) using a GE 3 Tesla HDx Excite MRI scanner. A Magnex head-dedicated gradient insert coil was used in conjunction with a birdcage, radiofrequency coil tuned to 127.7 MHz. Data were collected from 240 volumes each comprising 38 contigual axial slices via a gradient-echo EPI sequence (TR = 3 s, TE = 32 ms, FOV = 28.8 x 28.8 cm, matrix size = 128 x 128, voxel dimensions = 2.25 x 2.25 mm, slice thickness = 3 mm, flip angle = 90°). Visual stimuli were back-projected onto a custom in-bore acrylic screen at a distance of approximately 57 cm from the participant with images subtending approximately 9.5° of visual angle.

3.3.5 fMRI Analysis

Univariate analysis of the fMRI data was performed with FEAT v 5.98 (http://www.fmrib.ox.ac.uk/fsl). All analyses were performed separately for each experiment in the manner described below. In all scans the initial 9 s of data were removed to reduce the effects of magnetic stimulation. Motion correction (MCFLIRT, FSL) was applied followed by temporal high-pass filtering (Gaussian-weighted least-squares straight line fitting, sigma = 50s). Spatial smoothing (Gaussian) was applied at 6 mm (FWHM). Individual participant data were entered into a higher-level group analysis using a mixed-effects design (FLAME, http://www.fmrib.ox.ac.uk/fsl). Functional data were first registered to a high-resolution T1-anatomical image and then onto the standard MNI brain (ICBM152). A scene-selective region of interest was defined by the contrast of mixed>scrambled. The resulting group statistical maps were thresholded at Z>2.3. The thresholded statistical maps were then combined across experiments to generate a single scene-selective region of interest (ROI) used for subsequent MVPA analyses across both experiments (Figure A.1). We also generated a more restrictive ROI constrained to the scene-selective regions (parahippocampal place area (PPA), retrosplenial cortex (RSC) and the transverse occipital sulcus (TOS) or occipital place area (OPA)) that have been reported in previous fMRI studies (Epstein & Kanwisher, 1998; Maguire, 2001; Grill-Spector, 2003). This ROI was defined as follows; firstly, group mixed>scrambled statistical maps were averaged across the experiments. Next, seed points were defined at the peak voxels within this average statistical map for each region (PPA, RSC, TOS / OPA) in each hemisphere. The peak voxels of the ROIs had similar coordinates to those found in previous studies (Table A.1). For a given seed, a flood fill algorithm was used to identify a cluster of spatially contiguous voxels around that seed which exceeded a given threshold. This threshold was in turn iteratively adjusted till a cluster size of 500 voxels was achieved. This process was then repeated for each seed. Clusters for each region were combined across hemispheres to yield 3 ROIs each comprising 1000 voxels. Additionally, a single ROI combining all clusters across both hemispheres was defined. MNI co-ordinates of the seeds and corresponding thresholds are given in Table 3.1. All further analyses were restricted to these regions of interest.

Table 3.1. MNI mm co-ordinates and thresholds of standard place-selective (PPA, RSC, TOS / OPA) clusters.

Region	Hemisphere	*x*	*y*	*Z*	Threshold (Z)
PPA	L	-26	-48	-14	4.23
	R	30	-42	-16	4.24
RSC	L	-16	-60	4	3.58
	R	18	-56	6	3.77
TOS / OPA	L	-42	-84	20	3.52
	R	32	-88	12	3.28

Parameter estimates from the univariate analysis were normalised by subtracting the response to the mixed condition. Pattern analyses were then performed using the PyMVPA toolbox http://www.pymvpa.org/; Hanke et al. (2009). Figure 3.2 illustrates the method for determining the reliability of these neural patterns within and across subjects. To determine the reliability of the data within individual participants, the parameter estimates for each scene condition were correlated across odd (1, 3, 5, 7) and even (2, 4, 6, 8) blocks across all voxels in the scene-selective region (Haxby et al., 2001). The individual participant (IP) analysis was complemented by a group analysis, to determine the reliability of the pattern across participants. We used a leave-one-participant-out (LOPO) method (Shinkareva et al., 2008; Poldrack et al., 2009) in which the parameter estimates were determined using a group analysis of all participants except one. This generated parameter estimates for each scene condition in each voxel across the scene-selective region. This LOPO process was repeated such that every participant was left out of a group analysis once. For each LOPO iteration, the normalized patterns of response to each stimulus condition were correlated between the group and the participant that was left-out. This allowed us to determine whether there are reliable patterns of response that are consistent across individual participants. A Fisher's z-transformation was applied to the within-category and between-category correlations prior to further statistical analyses. For each category, the within-category and the average of the between-category correlations were calculated. These were entered into 3x2 repeated ANOVAs with the scene category (Experiment 1: city, indoor, natural; Experiment 2: coast, forest, mountain) and comparison (within, between) as the main factors. If neural response patterns to a given category can be distinguished from those to other categories, a significant main effect of comparison showing greater within- than between-category correlations would be expected. In order to obtain a measure of the decoding accuracy of our MVPA analyses, parameter estimates from the univariate analysis were also submitted to a k-nearest neighbour (kNN) classifier (k=1) using correlation as the distance measure.

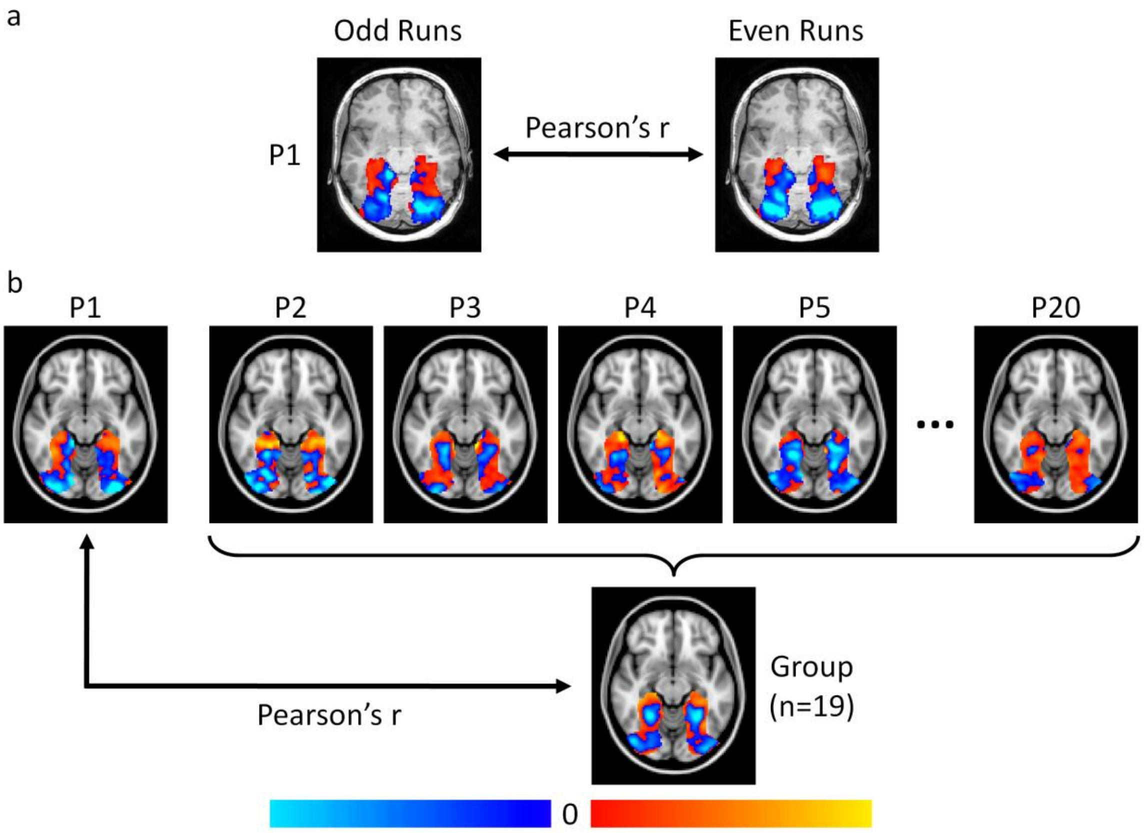


Figure 3.2. Schematic diagram of pattern analysis procedures. (a) Individual-participant (IP) analyses correlated neural patterns across odd and even runs of the stimulus presentation. (b) Group analyses compared individual patterns of response with the group pattern of response derived from all participants except that individual (LOPO). In both analyses this process is then repeated across all participants / LOPO iterations for all conditions.

In addition to the ROI analyses listed above we also performed whole-brain searchlight analyses (Kriegeskorte et al., 2006). A spherical ROI of radius 6mm was defined, and MVPA performed as described above. The average within- minus between-category correlation difference across categories was then assigned to the central voxel of the sphere, and the process repeated iterating the sphere over the whole-brain volume. A higher-level analysis using a mixed-effects design (FLAME) was used to determine whether the value at each voxel differed significantly from zero across individuals / LOPO-iterations. The resulting group statistical maps were thresholded at Z>2.3 with a cluster-correction of p<.05 applied.

3.3.6 Image Properties

Finally, we asked whether the patterns of neural response in Experiment 1 and 2 could be explained by the image statistics of the visual scenes. The image statistics of the scene images were computed using the GIST descriptor (http://people.csail.mit.edu/torralba/code/spatialenvelope/; Oliva and Torralba, 2001). First, each image is passed through a series of Gabor filters across 8 orientation and 4 spatial frequencies. This generates 32 filtered images. Next, each image is divided into a 4x4 grid giving 16 windows. The mean intensity is measured in each window. This generates a vector of 512 (32x16) values – the GIST descriptor – which represents the image in terms of the spatial frequencies and orientations present at different positions across the image. A schematic illustration of the calculation is given in Figure 3.3. In order to determine the similarity between individual scenes and the average of each scene category, GIST descriptors were correlated between each image and the average descriptor derived for each scene condition. This cross-validation procedure was used to determine how similar each image was to the average of its own category and to the other categories. Similarity with the neural response was determined by correlating the average GIST correlations matrix with the average MVPA correlations matrix. In order to assess the significance of this relationship, a simple regression analysis was performed using the average GIST correlations matrix as the regressor, and the corresponding MVPA correlation matrices concatenated across individuals / LOPO iterations as the outcomes. If the GIST correlations matrix is able to explain a significant amount of the variance in the corresponding MVPA correlation matrices, the model regression coefficient (Beta) can be expected to be significantly greater than zero. All regressor and outcome variables were

Z-scored prior to the regression analysis, such that all regression coefficients are given in standardised units. The image statistics of the scene images were also computed using pixelwise correlation of luminance values (cf Walther et. al, 2009). This provided us with a more basic image-based measure with which to compare with GIST descriptor.

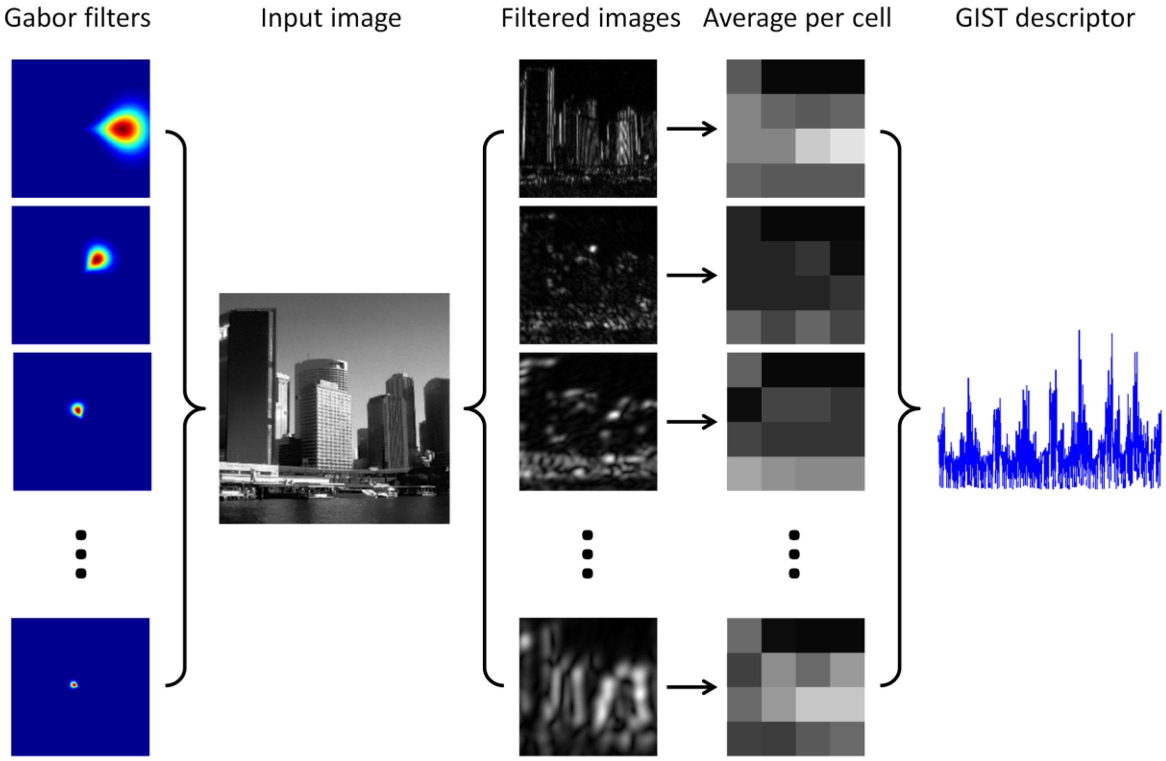


Figure 3.3. Schematic illustration of the calculation of a GIST descriptor for an example image. A series of Gabor filters across 8 orientations and 4 spatial frequencies are applied to the image. Each of the resulting 32 filtered images is then windowed by a 4x4 grid and the pixel intensities within each grid cell averaged together. Each grid cell thus represents the degree to which that window of the image is preserved by a Gabor filter at a given orientation and spatial frequency. The final GIST descriptor is a vector of 512 values yielded by concatenating these 16 cells across the 32 filtered images.

3.4 Results

3.4.1 Experiment 1

In the first experiment, we measured the patterns of response to different categories of visual scenes: city, indoor and natural. Figure 3.4 shows the normalized group response to city, indoor, and natural categories across the scene-selective region. Responses above the mean are shown in red and responses below the mean are shown in blue. Each category of scene had a distinct pattern of response, which was similar in appearance across the two cerebral hemispheres. Similar patterns were evident in individual participants (Figure A.2).

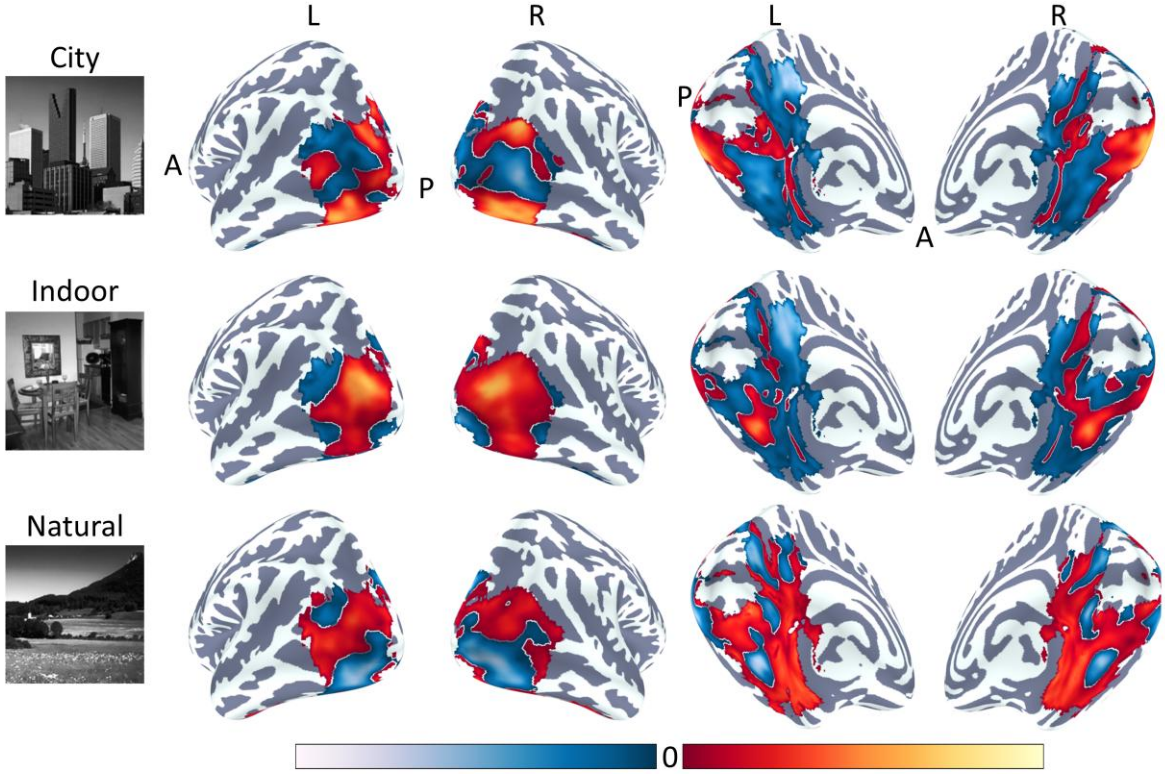


Figure 3.4. Experiment 1: Group patterns of response to city, indoor, and natural conditions on lateral (leftmost panels) and ventro-medial surfaces (rightmost panels). Patterns are restricted to regions defined by the response of mixed scenes > scrambled scenes. Red and blue colours indicate normalized values above and below the mean respectively.

Correlation based MVPA methods (Haxby et al., 2001) were used to measure the reliability of the neural response to these different categories of scene within individual participants (IP). Figure 3.5a shows a matrix of the correlations for the within- and between-category correlations. A 3 x 2 repeated measures ANOVA with Scene (city, indoor, natural) and Comparison (within, between) as the main factors showed a significant main effect of Comparison (F(1,19)=11.6, p=.003), showing that within-category correlations were higher than between-category correlations. However, there was no significant interaction between Scene * Comparison (F(2,38)=1.7, p=.196). A kNN classifier revealed that the decoding accuracy across categories was 46.7%, p=.008 (chance = 33%). A similar classification was evident when the ROI was restricted to all the standard scene-selective regions (combined PPA+RSC+TOS: 58.3%, p<.001). Figure A.3a shows the corresponding correlations matrix for this region. Splitting this ROI into its constituent regions revealed accuracies significantly above chance in PPA and TOS, but not RSC (PPA: 64.1%, p<.001; RSC: 42.5%, p=.09; TOS: 53.3%, p=.006).

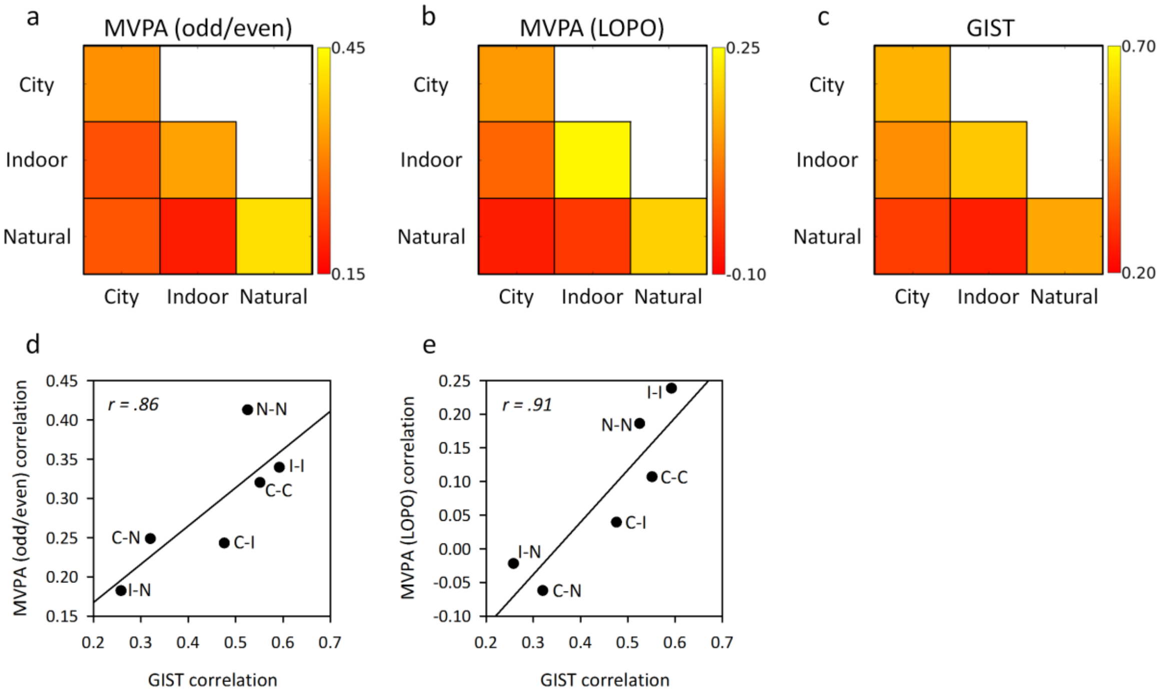


Figure 3.5. Experiment 1: Relationship between fMRI response and low-level image properties. Within- and between- category correlations for city, indoor, and natural conditions as determined by the individual-participant (a) and LOPO (b) MVPA analyses, and by the GIST image descriptor (c). Scatter-plots (d-e) showing strong positive correlations of the correlation matrices in (a) and (b) with (c) respectively.

We then determined the extent to which these patterns were consistent across participants using the LOPO method (see Methods). Figure 3.5b shows the correlation matrix using the LOPO method. There was a significant main effect of Comparison (F(1,19) = 90.8, p<.001), which was due to higher within-category compared to between-category correlations. There was also a significant Scene * Comparison interaction (F(2,38)=3.9, p=.028). This interaction was due to larger differences in within- versus between-category comparisons for the indoor and natural conditions compared to the city condition (city: p=.004, indoor: p<.001, natural: p<.001). A kNN classifier revealed a decoding accuracy across categories of 72.5%, p<.001 (chance = 33%). A similar classification was evident when the ROI was restricted to the standard scene-selective regions (combined PPA+RSC+TOS: 59.2%, p<.001); Figure A.3b shows the corresponding correlations matrix for this region. Splitting this ROI into its constituent regions revealed a similar pattern of results (PPA: 59.1%, p<.001; RSC: 50.8%, p=.003; TOS: 54.1%, p=.002).

To address the spatial scale of the patterns we repeated the LOPO and IP analyses with no spatial smoothing. Consistent with a coarser scale representation, we found a similar pattern of results (Figure A.4). We then repeated the LOPO and IP analyses using a whole-brain searchlight paradigm. Consistent with the previous analysis, we found that the majority of significant spheres clustered around the scene selective cortices defined by the ROI (Figure A.5).

Next, we used the GIST descriptor to measure the statistics of each image used in the fMRI experiment.   Figure 3.5c shows the within- and between-category correlations in image properties for different categories of visual scenes. We found higher within-category than between-category correlations (city: p<.001, indoor: p<.001, natural: p<.001). To determine whether there was a relationship between image properties of the stimuli and patterns of brain activity, the GIST correlations for each combination of scene were then correlated with the corresponding neural correlations for both the IP and LOPO. Figure 3.5d-e show the relationship between the similarity in image properties and the similarity in the pattern of response across different scenes. Strong positive correlations were evident for both the IP (r=.86) and LOPO analyses (r=.91). The significance of this relationship across participants or LOPO iterations was assessed using a simple regression analysis. The image properties significantly predicted the neural response in the IP (β=.28, p=.001) and LOPO analyses (β=.57, p<.001). A similar pattern of results was evident when the ROI was restricted to the standard scene-selective regions (combined PPA+RSC+TOS) for both the IP (r=.78, β=.32, p<.001) and LOPO analyses (r=.58, β=.33, p<.001); Figure A.3c-d. Splitting this ROI into its constituent regions produced a similar pattern of results for the IP analyses (PPA: r=.76, β=.46, p<.001; RSC: r=.75, β=.21, p=.022; TOS: r=.73, β=.21, p=.024) and LOPO analyses (PPA: r=.64, β=.42, p<.001; RSC: r=.55, β=.17, p=.065; TOS: r=.78, β=.26, p=.004).

We next repeated our analysis using pixel correlations as a measure of image properties. Pixel correlations did not significantly predict the neural response for the IP analysis (r=.12, β=.04, p=.653). However, a significant relationship was found for the LOPO analysis (r=.55, β=.34, p<.001). The pixel correlations were also poor predictors of the neural responses in the standard scene-selective regions for the IP analyses (combined PPA+RSC+TOS: r=.27, β=.11, p=.221; PPA: r=.01, β=.003, p=.973; RSC: r=.36, β=.10, p=.281; TOS: r=.31, β=.09, p=.339) and LOPO analyses (combined PPA+RSC+TOS: r=.17, β=.10, p=.298; PPA: r=.34, β=.12, p=.120; RSC: r=.24, β=.22, p=.017; TOS: r=.25, β=.08, p=.361). Thus, the pixel correlations measure was outperformed by the GIST descriptor.

3.4.2 Experiment 2

In the second experiment, we compared the patterns of responses to different types of natural landscapes: coasts, forests and mountains. Figure 3.6 shows the normalized group responses to coast, forest, and mountain scenes within scene-selective regions. Again, each category of scene had a distinct pattern of response, which was similar in appearance across the two cerebral hemispheres. Similar patterns of response can be found in the individual participants (Figure A.6). The reliability of these patterns of response was measured using the LOPO and IP methods. A 3 x 2 repeated measures ANOVA with Scene (coast, forest, mountain) and Comparison (within, between) as the main factors was used to test statistical significance.

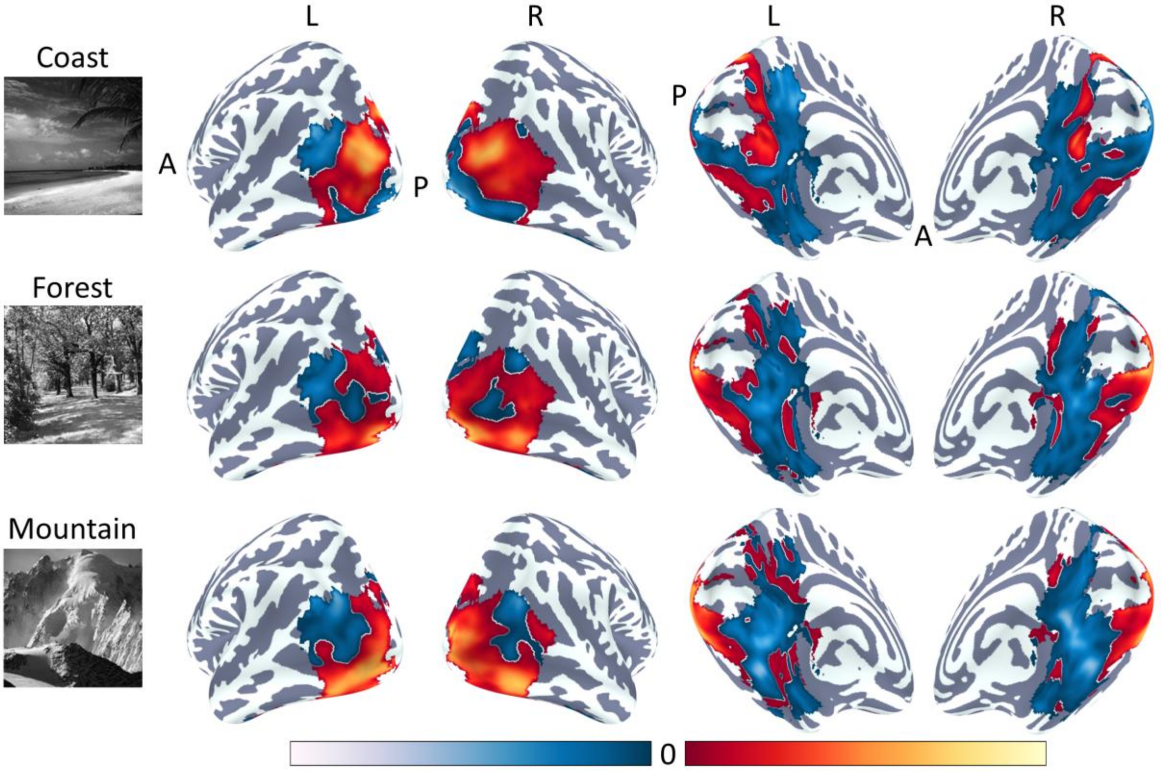


Figure 3.6. Experiment 2: Group patterns of response to coast, forest, and mountain conditions on lateral (leftmost panels) and ventro-medial surfaces (rightmost panels). Patterns are restricted to regions defined by the response of mixed scenes > scrambled scenes. Red and blue colours indicate normalized values above and below the mean respectively.

First, we performed the pattern analyses for individual participants (IP). The correlation between different scene categories is shown in Figure 3.7a. There was a significant main effect of Comparison (F(1,19)=33.30, p<.001), revealing significantly higher within-category compared to between-category correlations. However, there was not a significant Scene * Comparison interaction (F(2,38)=2.70, p=.079). A kNN classifier obtained mean decoding accuracy across all scene categories of 53.3%, p=.001 (chance=33%). A similar classification was evident when the ROI was restricted to the standard scene-selective regions (combined PPA+RSC+TOS: 55.8%, p<.001). Figure A.7a shows the corresponding correlations matrix for this region. Splitting this ROI into its constituent regions revealed accuracies significantly above chance in PPA and TOS, but not RSC (PPA: 56.7%, p<.001; RSC: 37.5%, p=.362; TOS: 52.5%, p=.002).

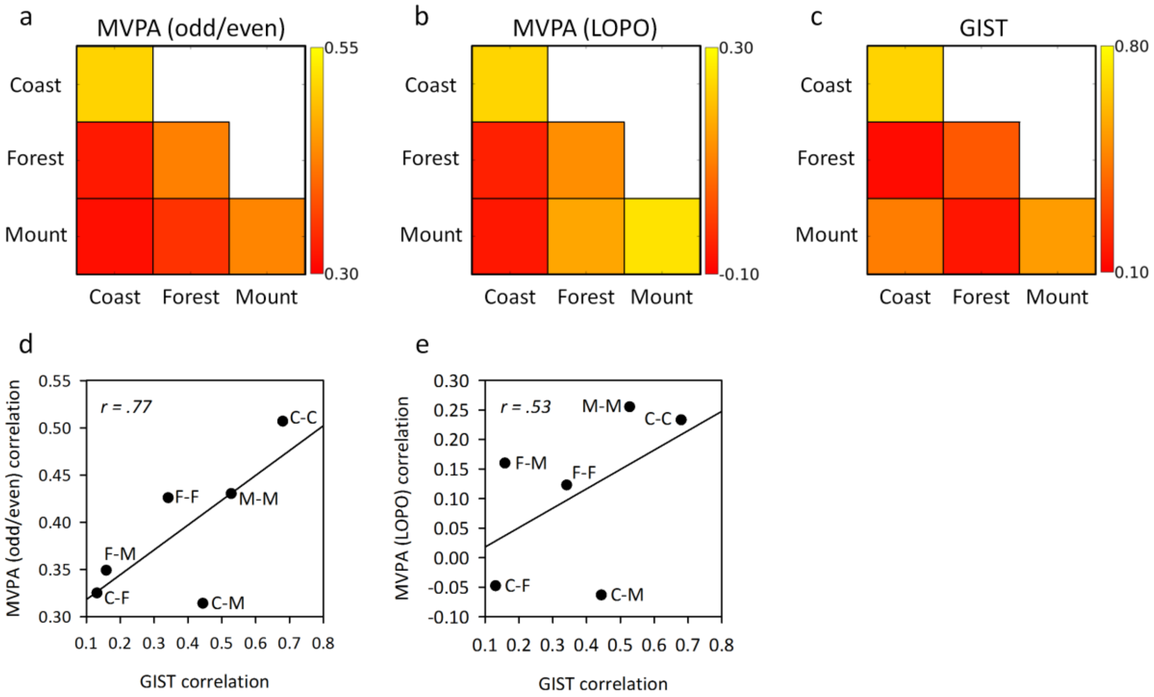


Figure 3.7. Experiment 2: Relationship between fMRI response and low-level image properties. Within- and between- category correlations for coast, forest, and mountain conditions as determined by the individual-participant (a) and LOPO (b) MVPA analyses, and by the GIST image descriptor (c). Scatter-plots (d-e) showing strong positive correlations of the correlation matrices in (a) and (b) with (c) respectively.

To determine whether the pattern of response was consistent across participants, we repeated the analysis using the LOPO method (Figure 3.7b). There was a significant main effect of Comparison (F(1,19)=114.40, p<.001) and a significant Scene * Comparison interaction (F(2,38)=18.18, p<.001). This interaction was due to larger within- versus between-category comparisons for the coast and mountain conditions compared to the forest condition (coast: p<.001, forest: p=.009, mountain: p<.001). A kNN classifier obtained mean decoding accuracy across all scene categories of 67.5%, p<.001 (chance=33%). A similar classification was evident when the ROI was restricted to the standard scene-selective regions (combined PPA+RSC+TOS: 50.8%, p<.001). Figure 3.7b shows the corresponding correlations matrix for this region. Splitting this ROI into its constituent regions revealed a similar pattern of results (PPA: 49.2%, p=.002; RSC: 50.0%, p=.002; TOS: 49.2%, p=.002).

To address the spatial scale of the patterns we repeated the LOPO and IP analyses with no spatial smoothing. Consistent with a coarser scale representation, we found a similar pattern of results (Figure A.8). To determine the extent to which our findings generalise to regions outside the ROI, the LOPO and IP analyses were repeated using a whole-brain searchlight paradigm. Significant spheres fell within the scene-selective ROI, particularly along the lateral regions, that included the TOS, and along medial regions that included the PPA and RSC. Figure A.9 shows the resulting searchlight group-average statistical maps.

Next, we used the GIST description to measure the statistics of each image used in the fMRI experiment. Figure 3.7c shows the within- and between-category correlations in image properties for different categories of visual scenes. We found higher within-category than between-category correlations (coast: p<.001, forest: p<.001, mountain: p<.001). To determine whether there was a relationship between image properties of the stimuli and patterns of brain activity, the GIST correlations for each combination of scene were then correlated with the corresponding neural correlations for both the IP and LOPO analyses. Figure 3.7d-e show the relationship between the similarity in image properties and the similarity in the pattern of response across different scenes for the IP and LOPO analyses. Positive correlations were evident for both the IP (r=.77) and LOPO analyses (r=.53). The significance of this relationship across participants / LOPO iterations was assessed using a simple regression analysis. The image properties significantly predicted the neural response in the IP (β=.27, p=.003) and LOPO analyses (β=.36, p<.001). A similar pattern of results was evident when the ROI was restricted to the standard scene-selective regions (combined PPA+RSC+TOS) for both the IP (r=.85, β=.27, p=.003) and LOPO analyses (r=.84, β=.37, p<.001); Figure A.7c-d. When the scene-selective ROI was split into its constituent regions, for the IP analysis the relationship between image properties and fMRI response was significant for the PPA and TOS, but not for the RSC (PPA: r=.70, β=.26, p=.004; RSC: r=.49, β=.08, p=.394; TOS: r=.91, β=.32, p<.001). The LOPO analysis showed a significant relationship for the TOS and RSC, but not in the PPA (PPA: r=.32, β=.13, p=.172; RSC: r=.56, β=.23, p=.012; TOS: r=.90, β=.24, p=.008).

We next repeated our analysis using pixel correlations as a measure of image properties. The pixel correlations significantly predicted the neural response for the IP (r=.70, β=.24, p=.008) but not the LOPO analyses (r=.23, β=.16, p=.084). When the ROI was restricted to the standard scene-selective regions, a more variable pattern of results was observed: IP analyses (combined PPA+RSC+TOS: r=.72, β=.22, p=.014; PPA: r=.59, β=.22, p=.014; RSC: r=.43, β=.07, p=.450; TOS: r=.76, β=.27, p=.003) and LOPO analyses (combined PPA+RSC+TOS: r=.69, β=.31, p=.001; PPA: r=.09, β= .04, p=.672; RSC: r=.38, β=.16, p=.088; TOS: r=.83, β=.22, p=.014). Although pixel correlations accounted for significant variance in the similarity of neural responses in some of the ROIs and analyses, performance was typically inferior to that of the GIST descriptor.

3.5 Discussion

The aim of this study was to understand the principles that underlie the organization of scene-selective regions of the human brain. We found that the patterns of response to images from the same scene category were more similar than the patterns of response from different categories of scene. However, there were differences in the magnitude of both the within- and between-category correlations. Next, we investigated the extent to which this variation in the categorical pattern of response to different scenes could be explained by systematic differences in image properties. We found a strong, linear relationship between the pattern of neural response in scene-selective regions and the image statistics of the scenes.

Our results show that the within-category correlations in fMRI responses to scenes were higher than the between-category correlations. These results are consistent with previous neuroimaging studies that have used pattern classification techniques to show distinct patterns of response to different categories of scene (Walther et al., 2009, 2011). However, our results also show that there was marked variation in the capacity of MVPA to distinguish different categories of real-world scenes. In Experiment 1, although we found distinct patterns of neural response to different categories of scenes, the patterns of response to natural landscapes were more distinct than to cityscapes or indoor scenes. In Experiment 2, we asked whether the patterns of response in scene-selective regions could discriminate between more subtle differences in scene type using different types of natural landscapes (coasts, forests, mountains). The results again showed that within-category responses were higher than between-category responses, but that there were also differences in the patterns of response to different types of natural scenes. For example, coastal scenes could be accurately distinguished from other scene categories on the basis of the pattern of brain activity they evoked, but the pattern of response to forests was often confused with the responses to mountain scenes.

The variability in the ability of the pattern of response to discriminate different scenes suggests that factors other than category membership may contribute to the organization of scene-selective regions. Other studies have found that classification of fMRI responses is impaired when poor exemplars of a scene are used (Torralbo et al., 2013). This suggests that the image properties may also be important. This conclusion is supported by other MVPA studies that have shown that variation in the pattern of response in scene-selective regions is not reflected by categorical differences in scenes, but rather by the spatial layout of the scene (Kravitz et al., 2011; Park et al., 2011). However, these studies do not provide a statistical account of how the spatial layout of the scene is linked to the patterns of response.

To directly address this issue, we determined the low-level properties of the images used in our experiment using the GIST descriptor (Oliva & Torralba, 2001). This determines the orientation energy at different spatial frequencies and spatial positions in the image and generates a list of values for each image that could be used to determine the similarity of images within and across different categories of scenes. The results showed that the properties of individual images of a scene were more similar to the average of images from the same category than they were to the average of images from different categories. However, like the neural patterns of response, there were also differences in the consistency or homogeneity of the image properties within different categories of scenes.

The main finding from this study was that the similarity of patterns of response to different categories of scenes showed a strong positive correlation with the similarity of their low-level image statistics. This relationship between the neural response and image properties was found in both experiments with two different methods of pattern analysis (IP, LOPO). The correlation is based not only on the variation within each category of scene, but also reflects systematic variation in the between-category confusions. Our findings contrast with those of Walther et al., (2009) who found no significant correlation between neural responses and image similarity. However, their analysis involved a different measure of image similarity based on correlating pixel values across images. Indeed, we likewise found that pixel correlations did not reliably predict the similarity of neural responses. The difference in results may reflect the fact that the GIST descriptor used in our main analysis more accurately reflects statistics encoded by the human visual system and was expressly devised to capture the critical spatial variables used to distinguish scene categories (Oliva & Torralba, 2001).

Whether we consider the ventral stream as a whole or whether we restrict our analysis to the standard scene-selective regions, the current findings suggest that the pattern of response to different categories of scenes is linked to the low-level properties of the image. This conclusion is consistent with other work showing that low-level image biases may be encoded in scene-selective regions. For example, spatial frequency (Rajimehr et al., 2011) and orientation (Nasr & Tootell, 2012) biases, along with visual field representations (Arcaro et al., 2009) have been reported in these regions.

Our results show that the neural patterns were not specific to individual participants; rather they reflect a more consistent functional organization. Using a modified cross-validation analysis (Haxby et al., 2001) we compared the pattern of response in one participant with the pattern from a group analysis in which that participant was left out. This leave-one-participant-out (LOPO) approach indicates that patterns of response to different visual scene categories are consistent across individuals (see also Shinkareva et al., 2008; Poldrack et al., 2009; Haxby et al., 2011). We found that the LOPO method often outperformed equivalent individual-participant (IP) analyses. These observations are significant in that they suggest that our findings reflect the operation of consistent, large-scale organizing principles, rather than an arbitrarily distributed representation in each individual.

In conclusion, our results showed that the pattern of response in scene-selective regions of the brain can be used to discriminate different categories of scene. However, there was systematic variation in the within- and between category similarity of neural responses across different scenes. We found that low-level image properties could explain these variations in response to visual scenes in scene-selective regions of the human brain.

Chapter 4 – Patterns of Neural Response in Scene-Selective Regions of the Human Brain are Affected by Low-Level Manipulations of Spatial Frequency

This chapter is adapted from: Watson, D. M., Hymers, M., Hartley, T., & Andrews, T. J. (2016). Patterns of neural response in scene-selective regions of the human brain are affected by low-level manipulations of spatial frequency. *Neuroimage*, *124*, 107–117. [2]

4.1 Abstract

Neuroimaging studies have found distinct patterns of response to different categories of scenes. However, the relative importance of low-level image properties in generating these response patterns is not fully understood. To address this issue, we directly manipulated the low level properties of scenes in a way that preserved the ability to perceive the category. We then measured the effect of these manipulations on category-selective patterns of fMRI response in the PPA, RSC, and OPA. In Experiment 1, a horizontal-pass or vertical-pass orientation filter was applied to images of indoor and natural scenes. The image filter did not have a large effect on the patterns of response. For example, vertical- and horizontal-pass filtered indoor images generated similar patterns of response. Similarly, vertical- and horizontal-pass filtered natural scenes generated similar patterns of response. In Experiment 2, low-pass or high-pass spatial frequency filters were applied to the images. We found that the image filter had a marked effect on the patterns of response in scene-selective regions. For example, low-pass indoor images generated similar patterns of response to low-pass natural images. The effect of filter varied across different scene-selective regions, suggesting differences in the way that scenes are represented in these regions. These results indicate that

[2] The author, David Watson, designed the experiment, analysed the results, and wrote the article under the supervision of Dr. Tom Hartley and Prof. Timothy Andrews. Mark Hymers provided technical assistance with the image filtering process and some statistical analyses.

patterns of response in scene-selective regions are sensitive to the low-level properties of the image, particularly the spatial frequency content.

4.2 Introduction

Despite their spatial complexity and heterogeneity, human observers are able to reliably categorise real world scenes even when images are presented rapidly (Potter, 1975; Greene & Oliva, 2009a) or visually degraded (Torralba, 2009; Walther et al., 2011). This capacity is thought to be based on neural activity in regions of human visual cortex that are selectively responsive to visual scenes (Aguirre & D'Esposito, 1997; Epstein & Kanwisher, 1998; Maguire, 2001; Nasr et al., 2011; Dilks et al., 2013). While studies using univariate fMRI analyses have reported comparable levels of response within these regions to different images of scenes (Epstein & Kanwisher, 1998), more recent reports employing multivariate techniques have shown that there are distinct patterns of response to different categories of scene (Walther et al., 2009, 2011) suggesting a finer-grained organisation that might underpin perceptual discriminations. However, the functional dimensions that shape these patterns have not been fully resolved.

Some reports have argued that patterns of response reflect high-level, categorical differences amongst scenes (Walther et al., 2009, 2011). For example, Walther and colleagues (2011) showed that the ability to decode scene categories from fMRI data was similar for photographs and line drawings, suggesting some level of invariance to the low level properties of images. However, other studies have suggested that patterns of response in scene-selective regions may be better explained in terms of visual properties of scenes such as spatial layout (e.g. Kravitz et al., 2011; Park et al., 2011; Watson et al., 2014). This latter account is consistent with the sensitivity of the amplitude of response in these regions for orientation (Nasr & Tootell, 2012), spatial frequency (Rajimehr et al., 2011; Musel et al., 2014), visual contrast (Kauffmann et al., 2015d), rectilinearity (Nasr et al., 2014), and visual field location (Levy et al., 2001; Arcaro et al., 2009; Golomb & Kanwisher, 2012). Nevertheless, these studies employed univariate analyses, so it remains unclear whether these modulations in the amplitude of response also affect the pattern of response.

In a recent study, we demonstrated that low-level properties of visual scenes, (defined by the GIST descriptor; Oliva and Torralba 2001), predicted patterns of neural response in scene-selective regions (Watson et al., 2014). However, images drawn from the same scene category are likely to have similar low-level properties (Oliva & Torralba, 2001). So, reliable category-specific patterns of response are expected under both categorical and image-based accounts. Therefore, it remains unclear whether patterns are determined primarily by membership of a common category or by the shared low-level image statistics characteristic of that category.

In the current study, we provide a direct comparison of the relative importance of image properties and category in determining patterns of response in scene-selective regions. Participants viewed images from two different categories of scene (indoor and natural) that are known to have distinct image properties (Oliva & Torralba, 2001) and to elicit different patterns of response in scene-selective regions (Walther et al., 2009; Watson et al., 2014). Low-level visual properties of the scenes were manipulated by filtering the images by orientation (Experiment 1) and spatial frequency (Experiment 2) as previous reports have suggested functional biases for these properties (Rajimehr et al., 2011; Nasr & Tootell, 2012). Using multi-voxel pattern analysis (MVPA), we compared the similarity of the patterns of neural response to each condition across the core scene regions (PPA, RSC, OPA). Our prediction was that if scene-selective regions are sensitive to image properties, then some degree of similarity should be seen between conditions sharing the same filter. If scene-selective regions are solely sensitive to category, then conditions sharing the same category should elicit similar patterns of response regardless of the low-level manipulation. The use of pattern analysis allows us to determine whether image properties are an important organizing factor in the topography of this region of the brain.

4.3 Methods

4.3.1 Participants

25 participants (8 males; mean age, 25.52; age standard deviation, 4.28; age range, 19-33) took part in Experiment 1 and 24 (8 males; mean age, 25.46; age standard deviation, 3.27;

age range, 20-32) took part in Experiment 2. All participants were neurologically healthy, right-handed, and had normal or corrected-to-normal vision. Written consent was obtained for all participants and the study was approved by the York Neuroimaging Centre Ethics Committee.

4.3.2 Stimuli

Visual stimuli were back-projected onto a custom in-bore acrylic screen at a distance of approximately 57 cm from the participant with all images subtending approximately 10.7° of visual angle. Images presented in the main experiment runs were taken from the LabelMe scene database (http://cvcl.mit.edu/database.htm; Oliva & Torralba, 2001) and presented in greyscale. The image set comprised 128 images; 64 indoor and 64 natural scenes. These categories were selected on the basis of their inclusion in previous studies of scene processing (Oliva & Torralba, 2001; Walther et al., 2009). Images were first converted to greyscale – this is important as the filtering process can produce undesirable artifacts in colour images. For instance, high-pass filtering a colour image is likely to introduce false colour into areas of the image not passed by the filter, which will now appear a colour given by the mean luminance of each colour channel. Next, luminance histograms were equated across all images using the MATLAB SHINE toolbox (Willenbockel et al., 2010) prior to any filtering. The full sets of unfiltered indoor and natural images are shown in Figures A.10 and A.11 respectively.

Filtering was performed by weighting the Fourier spectrum of each image to preserve either horizontal or vertical orientations (Experiment 1), or high or low spatial frequencies (Experiment 2). In Experiment 1, filters were wrapped Gaussian profiles, with a wide angle cut-off (FWHM = 75°) that ensured images remained recognisable after filtering. In Experiment 2 filters were Gaussian profiles with cut-offs set at less than 2 cycles/degree and greater than 6 cycles/degree at FWHM for the low- and high-pass filters respectively. Filter cut-offs for Experiment 2 were based upon those used in previous literature (Schyns & Oliva, 1994, 1999; Oliva & Schyns, 1997). A soft window was applied around the edges of all images to reduce wrap-around edge artifacts associated with the filtering process. Figure 4.1 shows examples of the images used in each experiment.

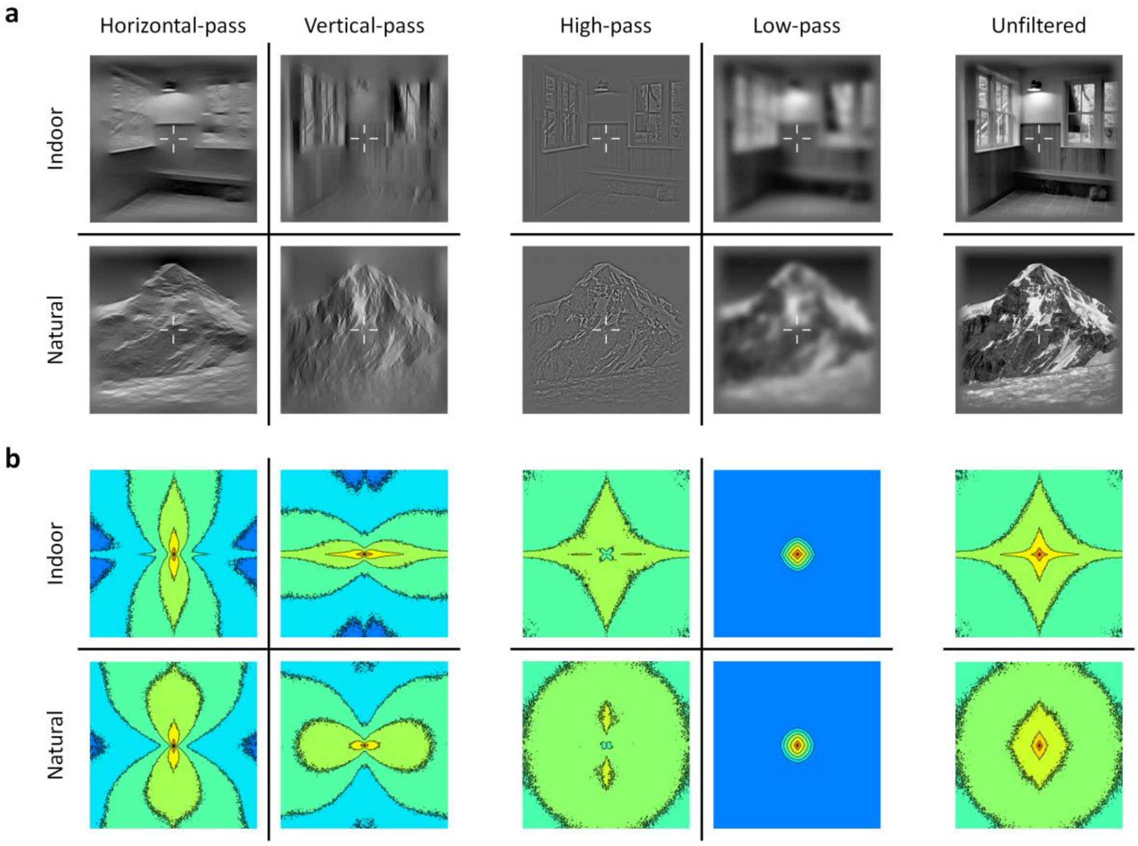


Figure 4.1. (a) Examples of images from conditions in Experiment 1 (left panel) and Experiment 2 (middle panel). For comparison, equivalent unfiltered images are shown (right panel). (b) Average Fourier amplitude spectra across all images in each condition.

For each experiment, an additional localiser scan was performed. An independent set of 64 scene images were drawn from the SUN database (http://groups.csail.mit.edu/vision/SUN/; Xiao et al., 2010) and presented in full colour. The SUN database is hierarchically organised into manmade-indoor, manmade-outdoor and natural-outdoor scenes, and stimuli were drawn in approximately equal numbers from each of these 3 classifications. Fourier-scrambled images were created by applying the same set of random phases to each 2-dimensional frequency component in each colour channel of the original image while keeping the magnitude constant. Intact and scrambled images were then rescaled to have a mean luminance equal to that of the images used in the experimental scan. Figure 4.2a shows examples of the images used in the localiser scan.

4.3.3 Experimental Design

During the localiser scan, participants viewed images from 2 stimulus conditions: 1) intact scene images and 2) phase scrambled versions of the same images in condition 1. During the experimental scan participants viewed images from 4 stimulus conditions comprising 2 scene categories (indoor and natural) across 2 levels of filtering (Experiment 1: horizontal-pass, vertical-pass; Experiment 2: low-pass, high-pass). Stimuli were presented using PsychoPy (Peirce, 2007, 2009).

In both the localiser and experimental scans, images from each condition were presented in a blocked fMRI design with 9 images per block (8 unique and 1 repeated). Each image was presented for 750ms followed by a 250ms grey screen that was equal in mean luminance to the scene images. Each stimulus block was separated by a 9s period in which the same grey screen as used in the inter-stimulus interval was presented. In order to minimise eye movements a central fixation cross was superimposed on all images and the grey screen and participants were instructed to maintain fixation for the duration of both scans. Each condition was repeated 8 times in a counterbalanced block design giving a total of 16 and 32 blocks in the localiser and experimental scans respectively. To maintain attention throughout the scan sessions participants performed a one-back task in which they were required to detect the repeated presentation of one image in each block, responding to the repeated image with a button press. By using a passive task we avoid biasing neural responses towards either one of our experimental manipulation; for instance, a categorisation task might bias responses towards the category manipulation, whereas an image-based task might bias responses towards the filter manipulation.

4.3.4 Imaging Parameters

All scanning was conducted at the York Neuroimaging Centre (YNiC) using a GE 3 Tesla HDx Excite MRI scanner. A Magnex head-dedicated gradient insert coil was used in conjunction with a birdcage, radiofrequency coil tuned to 127.7MHz. Data were collected from 38 contigual axial slices via a gradient-echo EPI sequence (TR = 3s, TE = 32.5ms, FOV = 288x288mm, matrix size = 128x128, voxel dimensions = 2.25x2.25 mm, slice thickness = 3mm, flip angle = 90°).

4.3.5 fMRI Analysis

Univariate analyses of the fMRI data were performed with FEAT v5.98 (http://www.fmrib.ox.ac.uk/fsl). In all scans the initial 9s of data were removed to reduce the effects of magnetic stimulation. Motion correction (MCFLIRT, FSL, Jenkinson et al., 2002) was applied followed by temporal high-pass filtering (Gaussian-weighted least-squares straight line fittings, sigma=50s). Spatial smoothing (Gaussian) was applied at 6mm FWHM to both the localiser and experiment runs, in line with previous studies employing smoothing in conjunction with MVPA (Op de Beeck, 2010; Watson et al., 2014). Parameter estimates were generated for each condition by regressing the hemodynamic response of each voxel against a box-car regressor convolved with a single-gamma HRF. Next, individual participant data were entered into higher-level group analyses using a mixed-effects design (FLAME, FSL). Functional data were first registered to a high-resolution T1-anatomical image and then onto the standard MNI brain (ICBM152).

A scene-selective region of interest was defined from the localiser data of both experiments using the contrast of intact scenes > scrambled scenes (Figure 4.2b). The intact scenes share the same amplitude spectra with their phase scrambled counterparts, thus such a contrast provides a clearer control for low-level visual differences than other commonly used contrasts such as scenes > objects or scenes > faces. For instance, although scenes and objects / faces differ in their category membership, they also differ in a large number of image properties (e.g. spatial frequency, orientation, retinotopic eccentricity, etc.). Given that this experiment aimed to investigate the neural representation of image properties, it was important to use the contrast that provided a stronger control for such visual differences. This ROI therefore provides a definition including scene-selective voxels across a wide extent of cortex – this enables us to test the distributed neural representations of the images as originally described by Haxby et al. (2001). This scene-selective ROI was used for subsequent MVPA across both experiments.

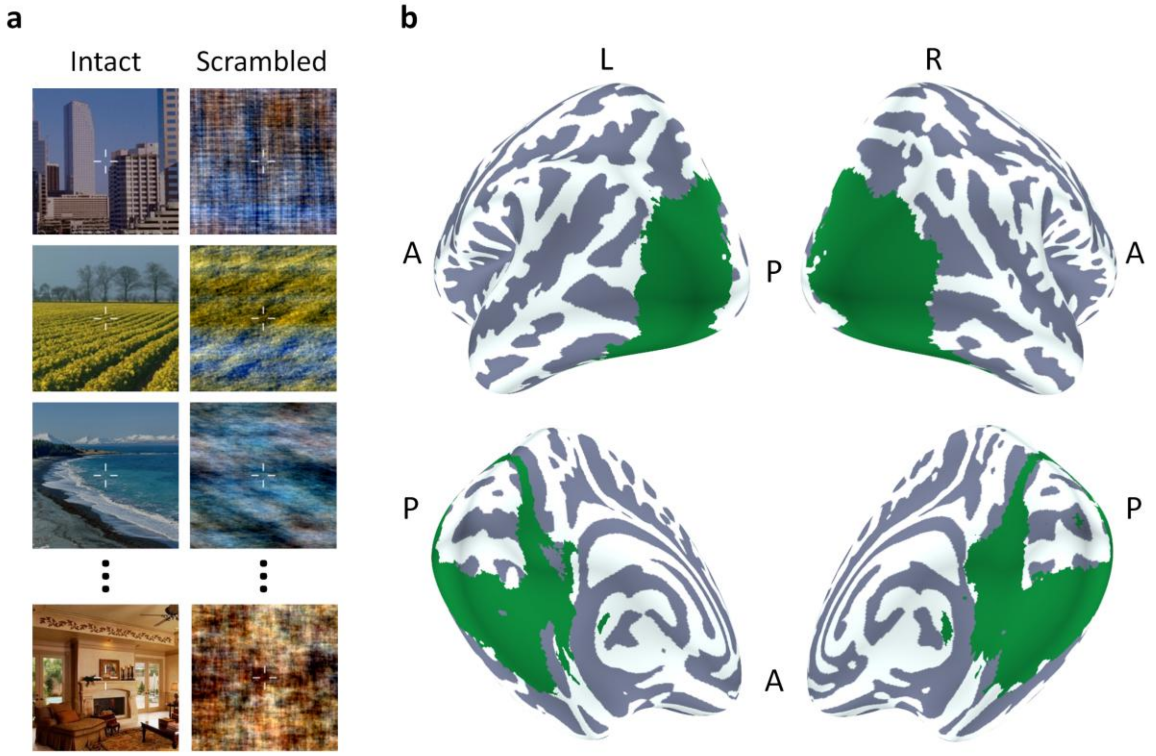


Figure 4.2. (a) Examples of images presented in the localiser scan. (b) Mask used for ROI analyses defined by the contrast of intact > scrambled.

We also generated more restrictive ROIs constrained to the classical scene-selective regions (parahippocampal place area (PPA), retrosplenial complex (RSC), occipital place area (OPA)) that have been reported in previous fMRI studies (Epstein & Kanwisher, 1998; Maguire, 2001; Dilks et al., 2013). Within the MNI-2x2x2mm space, group intact>scrambled statistical maps were first averaged across the experiments. Next, seed points were defined at the peak voxels within the average intact>scrambled statistical map for each region (PPA, RSC, OPA) in each hemisphere. For a given seed, a flood fill algorithm was used to identify a cluster of spatially contiguous voxels around that seed which exceeded a given threshold. This threshold was then iteratively adjusted till a cluster size of approximately 500 voxels was achieved (corresponding to a volume of 4000mm^3); actual cluster sizes ranged from 499-501 voxels as an optimal solution to the algorithm was not always achievable. This step ensures that estimates of multi-voxel pattern similarity are not biased by the different sizes of ROIs being compared. Clusters were combined across hemispheres to yield 3 ROIs, each comprising approximately 1000 voxels. These regions are shown in Figure 4.3. MNI co-ordinates of the seeds are given in Table 4.1. These seed points had similar locations to those reported in previous literature (Table A.1). To ensure clusters were appropriately sized we additionally repeated our analyses across using clusters across a range of sizes from 200-500 voxels. We found that the cluster size made little to no difference upon the main results (Figure A.12). An additional early visual control ROI was defined from the V1 region of the Jülich histological atlas (Amunts et al., 2000; Eickhoff et al., 2005). We also tested for possible differences in response within the PPA region by splitting this region precisely halfway along its posterior-anterior extent into a posterior PPA and an anterior PPA region.

Table 4.1. Peak MNI mm co-ordinates and thresholds of standard scene-selective clusters (PPA, RSC, OPA).

Region	Hemisphere	*x*	*y*	*z*	Threshold (Z)
PPA	L	-24	-52	-14	5.21
	R	26	-50	-16	5.68
RSC	L	-18	-62	4	4.24
	R	16	-54	-2	4.92
OPA	L	-36	-88	4	5.23
	R	36	-82	4	5.54

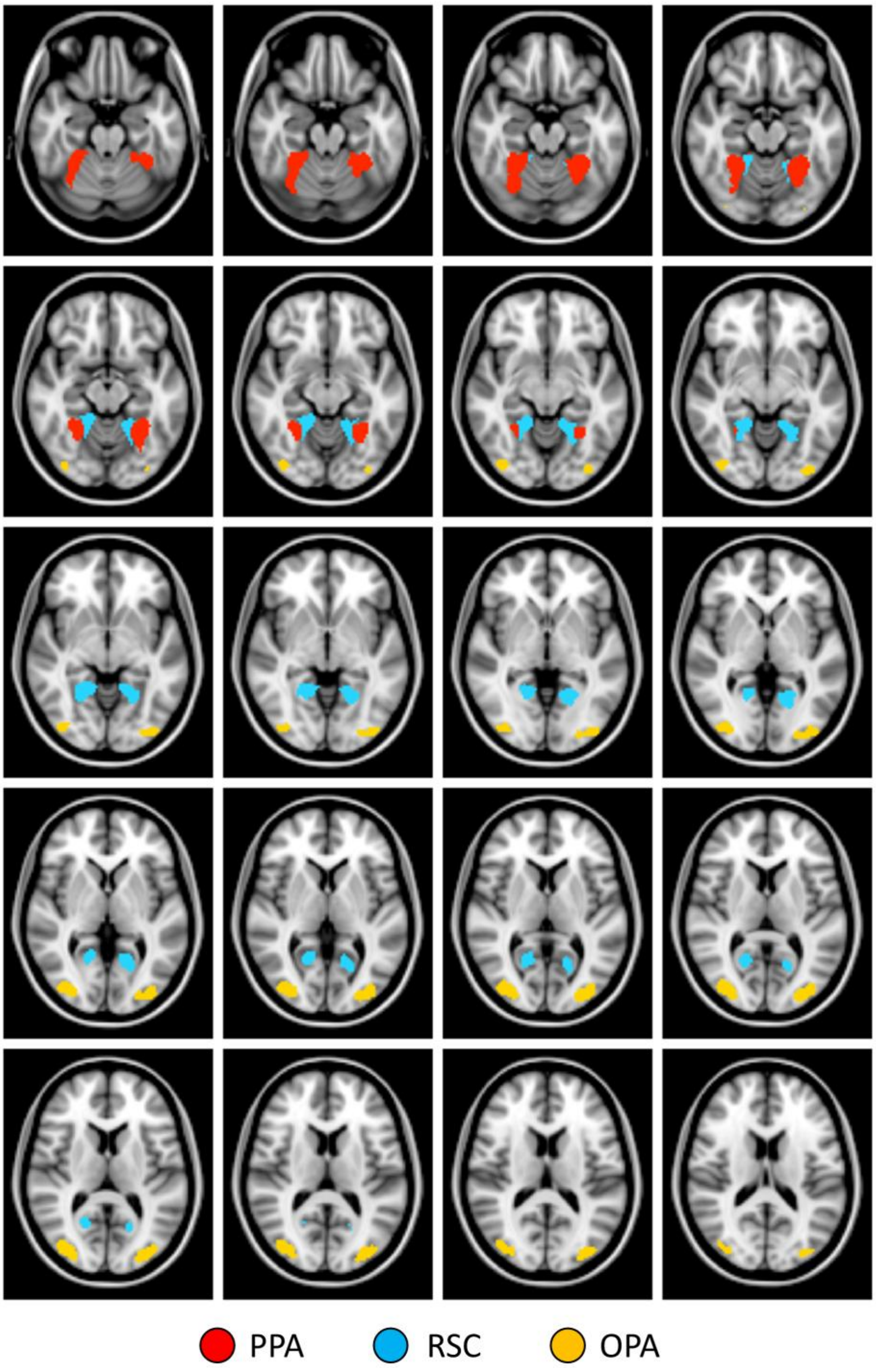


Figure 4.3. Masks used for ROI analyses of core scene regions. Each mask comprises approximately 500 voxels (4000mm^3) in each hemisphere. Slices of MNI brain span the range from Z = -22 to Z = 16 in 2mm increments.

Next, we measured patterns of response to different stimulus conditions in each experiment. Parameter estimates were generated for each condition in the experimental scans. The reliability of response patterns was tested using a leave-one-participant-out (LOPO) cross-validation paradigm (Shinkareva et al., 2008; Poldrack et al., 2009) in which parameter estimates were determined using a group analysis of all participants except one (Figure 3.2). This generated parameter estimates for each scene condition in each voxel. This LOPO process was repeated such that every participant was left out of a group analysis once. These data were then submitted to correlation-based pattern analyses (Haxby et al., 2001, 2014) implemented using the PyMVPA toolbox (http://www.pymvpa.org/; Hanke et al., 2009). Parameter estimates were normalised by subtracting the mean response per voxel across all experimental conditions (see Haxby et al., 2001). For each iteration of the LOPO cross-validation, the normalized patterns of response to each stimulus condition were correlated between the group and the left-out participant. This allowed us to determine whether there are reliable patterns of response that are consistent across individual participants. A Fisher's z-transformation was then applied to the correlations prior to further statistical analyses.

We next used a representational similarity analysis (RSA; Kriegeskorte et al. 2008) utilising multiple regression to assess the relative contributions of category information and image properties to the neural response patterns. For each factor (category and filter type) a binary regressor was generated representing a model correlations matrix whereby ones were placed on those elements where the relevant factor was shared and zeroes on all other elements. The regressors therefore represent the extreme cases where the patterns of response are entirely predicted by either the scene category or by the filtering; these regressors are illustrated for Experiments 1 and 2 in Figure 4.5a-b and Figure 4.8a-b respectively. Each regressor was then repeated and tiled across LOPO iterations. The outcomes measure was defined as the MVPA correlation matrices concatenated across LOPO iterations. All regressors and outcomes were then Z-scored such that all outputs of the regression model are given in standardised units. These regressors and outcomes were then entered into the multiple regression model. This analysis yielded a beta value and associated standard error for each regressor which would be expected to differ significantly from zero if that regressor were able to explain a significant amount of the variance in the MVPA correlations. A t-contrast was used to assess the significance of the differences between the betas.

4.3.6 Behavioural Experiment

In order to ensure that the filtering process did not disrupt the ability of participants to perceive the scenes categorically, we conducted an additional behavioural experiment. A new set of 20 participants (5 males; mean age, 26.80; age standard deviation, 3.32; age range, 23-34) were presented with the images used in the fMRI experiments plus their unfiltered equivalents. This produced 10 conditions across 2 categories (indoor, natural) and 5 levels of filtering (horizontal-pass, vertical-pass, high-pass, low-pass, unfiltered). For each participant, images were divided into 5 subsets and then each subset randomly assigned to a different filtering condition such that participants only saw each image once across all filtering conditions. A chin rest was used to maintain viewing distance across participants. Images subtended a visual angle of approximately 10.7°. In each trial a fixation screen was presented for 1000ms, followed by an image for 750ms. Importantly, both visual angle and stimulus duration were set to match those of the fMRI experiment. Following this, a blank screen was presented for 2250ms or until the participant made a response. Participants were required to indicate, with a button press, whether the image was of an indoor or natural scene as quickly and as accurately as possible, and were able to respond immediately after stimulus onset.

4.4 Results

4.4.1 Experiment 1

In Experiment 1, we measured patterns of neural response to different categories of scene (indoor and natural) filtered by orientation (horizontal-pass and vertical-pass). Figure 4.4 shows the normalised group responses to each condition across the scene-selective ROI. Responses above the mean are shown in red and responses below the mean are shown in blue.

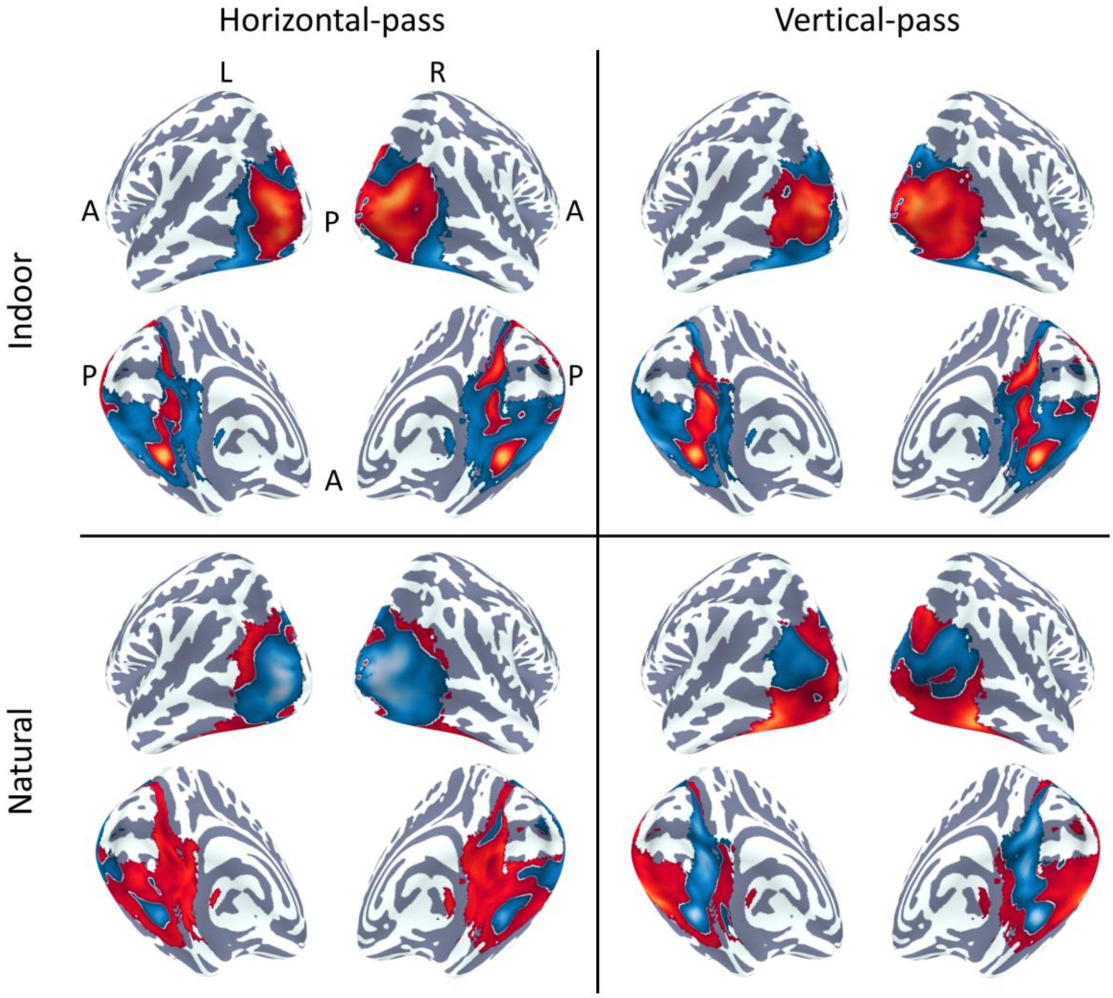


Figure 4.4. Group patterns of response to conditions in Experiment 1. Patterns are restricted to regions defined by the response of intact scenes > scrambled scenes. Red and blue colours indicate normalized values above and below the mean respectively.

A correlation based MVPA (Haxby et al., 2001) was conducted to measure the similarity of the neural responses to different conditions (Figure 4.5c). To test the contribution of category and image factors to the neural responses, we used a representational similarity analysis (Kriegeskorte et al., 2008). Model correlation matrices were generated representing the extreme cases where the patterns of response are entirely predicted by the scene category (Figure 4.5a) or by the orientation filter (Figure 4.5b). These were then used as regressors in a multiple regression analysis of the fMRI data. Figure 4.5d shows the resulting coefficients for each regressor. Both the category (β = 0.82, p < .001) and filter regressors (β = 0.17, p < .001) explained a significant amount of the variance in the MVPA correlation matrix. However, a t-contrast revealed that the category regressor explained significantly more variance than the filter regressor (t = 12.84, p < .001). A series of post-hoc paired-sample t-tests were used to compare the critical elements of the correlations matrix representing the same-category, different-filter and different-category, same-filter correlations. In all cases, the same-category/different-filter correlations were found to be significantly greater than the different-category/same-filter correlations (indoor-horizontal-pass/indoor-vertical-pass > indoor-horizontal-pass/natural-horizontal-pass: t(24) = 13.32, p < .001; natural-horizontal-pass/natural-vertical-pass > indoor-horizontal-pass/natural-horizontal-pass: t(24) = 7.07, p < .001; indoor-horizontal-pass/indoor-vertical-pass > indoor-vertical-pass/natural-vertical-pass: t(24) = 14.68, p < .001; natural-horizontal-pass/natural-vertical-pass > indoor-vertical-pass/natural-vertical-pass: t(24) = 8.64, p < .001). An additional post-hoc test did not find a significant difference between correlations in the indoor-horizontal-pass/natural-horizontal-pass and the indoor-vertical-pass/natural-vertical-pass comparison (t(24) = 1.13, p = .271). Thus, patterns were no more or less similar for horizontal-pass than vertical-pass filtered images.

Restricting the regression analysis to the standard scene-selective regions (PPA, RSC, OPA) revealed a similar pattern of results (Figure 4.6). Responses in the PPA were significantly predicted by the category (β = 0.85, p < .001) but not the filter regressor (β = 0.04, p = .204), with significantly more variance explained by the category than the filter regressor (t = 16.34, p < .001). Responses in the RSC were significantly predicted by the category (β = 0.77, p < .001) but not the filter regressor (β = 0.02, p = .529), with significantly more variance explained by the category than the filter regressor (t = 12.01, p < .001). Responses in the OPA were significantly predicted by the category (β = 0.73, p < .001) but not the filter regressor (β = 0.07, p = .095), with significantly more variance explained by the category than the filter regressor (t = 10.21, p < .001). In contrast to the scene regions, responses in the early visual (V1) control region were significantly predicted by both the category (β = 0.36, p < .001) and filter regressors (β = 0.25, p < .001). There was no significant difference between the effect of category and filter (t = 1.28, p = .203). Results of post-hoc t-tests for these regions are given in Table 4.2.

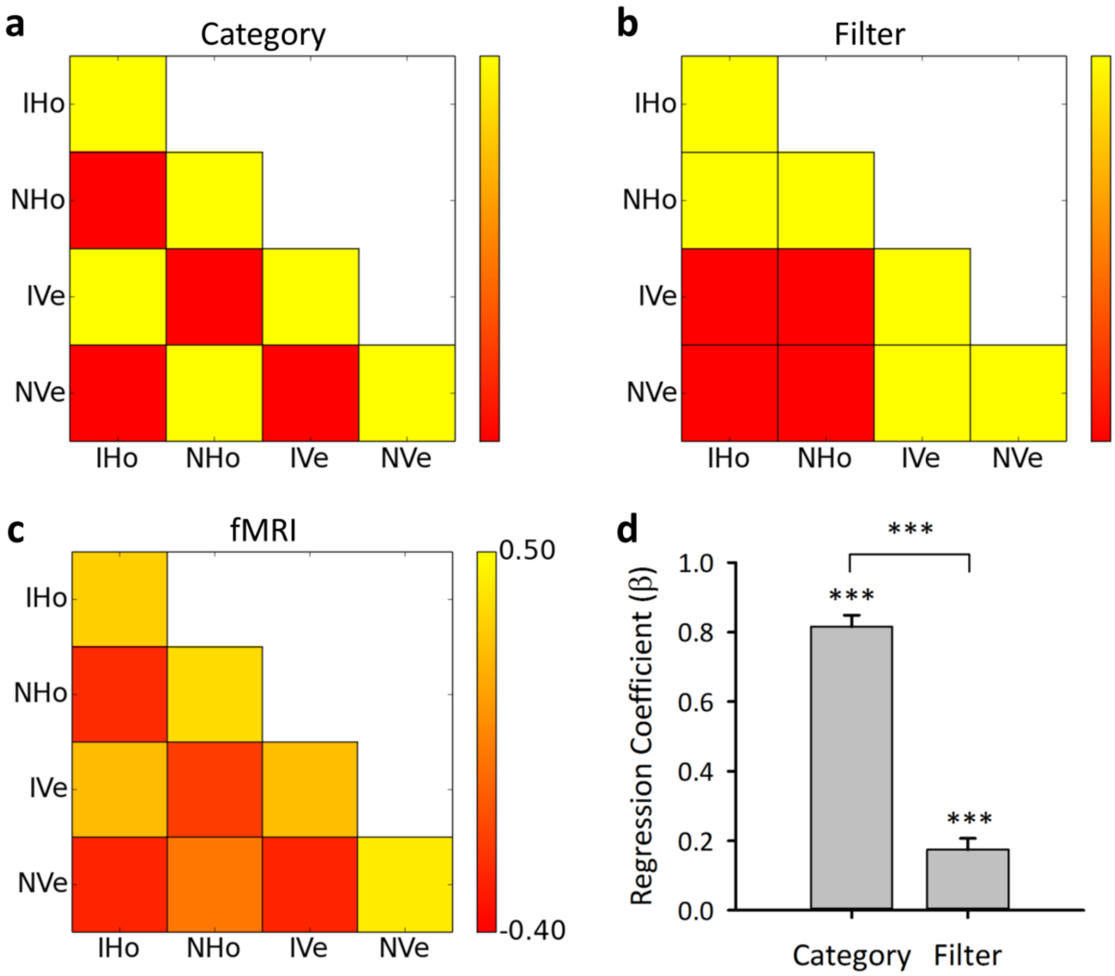


Figure 4.5. Experiment 1 analysis. Condition labels: indoor horizontal-pass (IHo), natural horizontal-pass (NHo), indoor vertical-pass (IVe), natural vertical-pass (NVe). Binary models were defined representing the cases where the patterns of response are entirely predicted by either the category (a) or the filter type (b). These were entered into a multiple regression analysis as regressors, while the fMRI MVPA correlations (c) were entered as outcomes. The resulting regression coefficients are shown in (d). Error bars represent 1 SEM. (* p < .05, ** p < .01, *** p < .001).

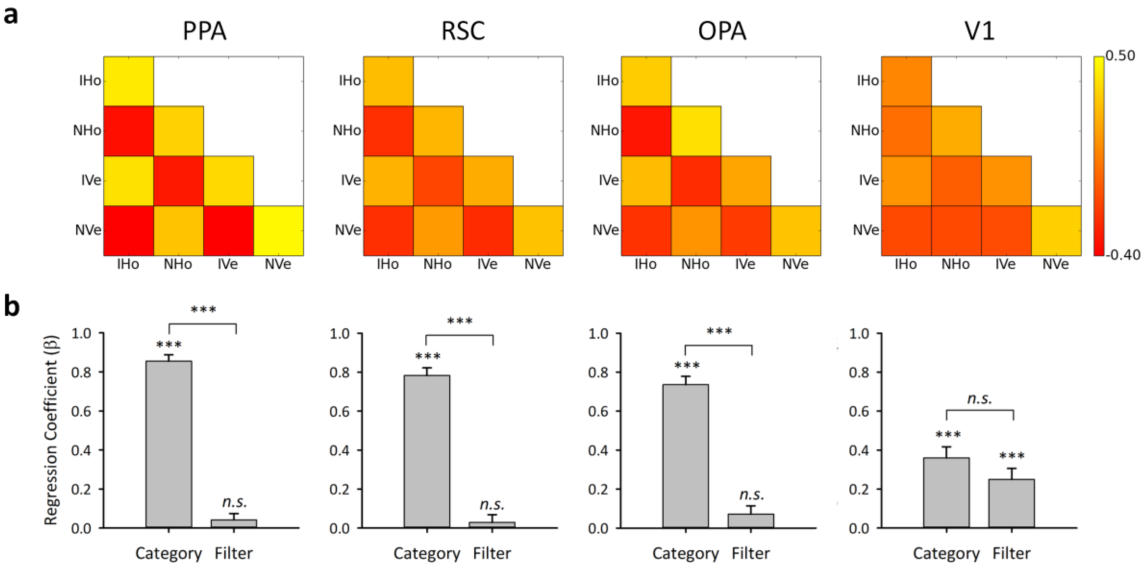


Figure 4.6. Experiment 1: standard scene-selective regions and V1. (a) MVPA correlation matrices. (b) These matrices were compared against binary regressors of category and filter effects using a multiple regression analysis; resulting beta coefficients are shown for each regressor. Error bars represent 1 SEM. (* p < .05, ** p < .01, *** p < .001).

Table 4.2. Experiment 1: t-statistics and significance of post-hoc pairwise t-tests for standard scene selective regions (PPA, RSC, OPA) and V1. (* p < .05, ** p < .01, *** p < .001)

	PPA	RSC	OPA	V1
IHo/IVe > IHo/NHo	9.35***	9.30***	9.25***	2.49(ns)
NHo/NVe > IHo/NHo	8.68***	8.77***	7.05***	-2.13(ns)
IHo/IVe > IVe/NVe	9.84***	7.14***	6.97***	3.83**
NHo/NVe > IVe/NVe	9.26***	7.05***	5.09***	-0.18(ns)

4.4.2 Experiment 2

In Experiment 2, we measured patterns of neural response to different categories of scene (indoor and natural) filtered by spatial frequency (high-pass and low-pass). Figure 4.7 shows the normalised group responses to each condition across the scene-selective ROI. Responses above the mean are shown in red and responses below the mean are shown in blue.

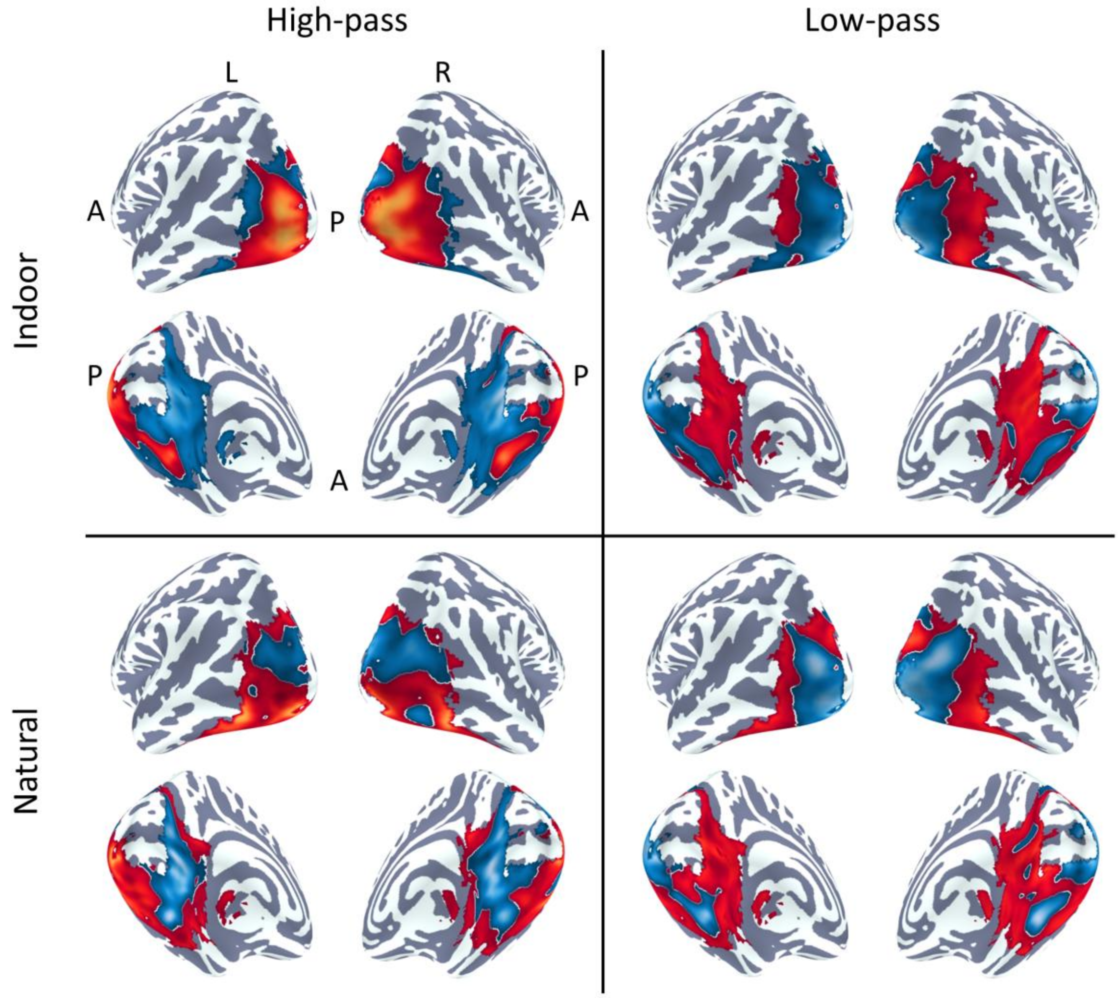


Figure 4.7. Group patterns of response to conditions in Experiment 2. Patterns are restricted to regions defined by the response of intact scenes > scrambled scenes. Red and blue colours indicate normalized values above and below the mean respectively.

Correlation based MVPA was used to assess the similarity of the neural responses across different conditions. The influence of category and image factors on the fMRI data was assessed using a representational similarity analysis. Model correlation matrices representing the cases where responses are entirely predicted by the scene category (Figure 4.8a) or by the spatial frequency filtering (Figure 4.8b) were entered as regressors in a multiple regression analysis of the fMRI data (Figure 4.8c). Figure 4.8d shows the resulting coefficients for each regressor. Both the category (β = 0.23, p < .001) and filter regressors (β = 0.86, p < .001) explained a significant amount of the variance in the MVPA data. However, in contrast to Experiment 1, the filter regressor explained significantly more variance than the category regressor (t = 16.93, p < .001). Post-hoc tests revealed greater different-category/same-filter than same-category/different-filter correlations in all cases (indoor-high-pass/natural-high-pass > indoor-high-pass/indoor-low-pass: t(23) = 17.56, p < .001; indoor-high-pass/natural-high-pass > natural-high-pass/natural-low-pass: t(23) = 10.29, p < .001; indoor-low-pass/natural-low-pass > indoor-high-pass/indoor-low-pass: t(23) = 20.26, p < .001; indoor-low-pass/natural-low-pass > natural-high-pass/natural-low-pass: t(23) = 15.95, p < .001). An additional post-hoc test revealed significantly higher correlations in the indoor-low-pass/natural-low-pass than the indoor-high-pass/natural-high-pass comparison (t(23) = 10.51, p < .001), indicating greater similarity in the neural response patterns across low-pass than high-pass filtered images.

Restricting the regression analyses to the standard scene-selective regions (PPA, RSC, OPA) revealed a more variable pattern of results (Figure 4.9). Responses in the PPA were significantly predicted by both the category (β = 0.66, p < .001) and filter regressors (β = 0.43, p < .001). However, in contrast to the scene-selective region as a whole, more variance was explained by the category than the filter regressor (t = 4.33, p < .001) in this subregion. Responses in the RSC were significantly predicted by both the category (β = 0.35, p < .001) and filter regressors (β = 0.53, p < .001) but in this case slightly more variance was explained by the filter than the category regressor (t = 2.41, p = .017). Responses in the OPA were significantly predicted by both the category (β = 0.22, p < .001) and filter regressors (β = 0.66, p < .001), but again significantly more variance was explained by the filter than the category regressor (t = 6.25, p < .001). Responses in the V1 control region were significantly predicted by the filter (β = 0.95, p < .001) but not the category regressor (β = 0.03, p = .213), with significantly more variance explained by the filter than the category regressor (t = 29.96, p < .001). Results of post-hoc t-tests for these regions are given in Table 4.3.

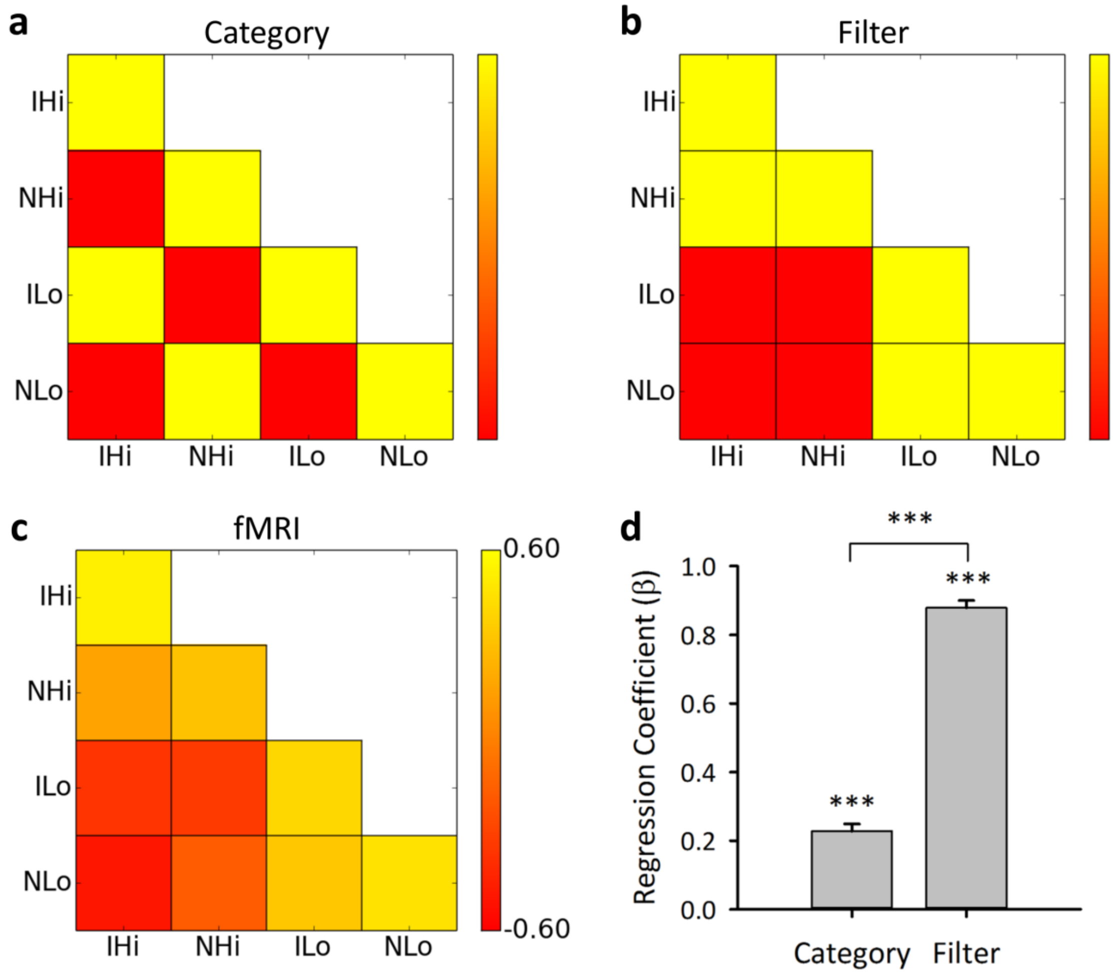


Figure 4.8. Experiment 2 analysis. Condition labels: indoor high-pass (IHi), natural high-pass (NHi), indoor low-pass (ILo), natural low-pass (NLo). Binary models were defined representing the cases where the patterns of response are entirely predicted by either the category (a) or the filter type (b). These were entered into a multiple regression analysis as regressors, while the fMRI MVPA correlations (c) were entered as outcomes. The resulting regression coefficients are shown in (d). Error bars represent 1 SEM. (* p < .05, ** p < .01, *** p < .001).

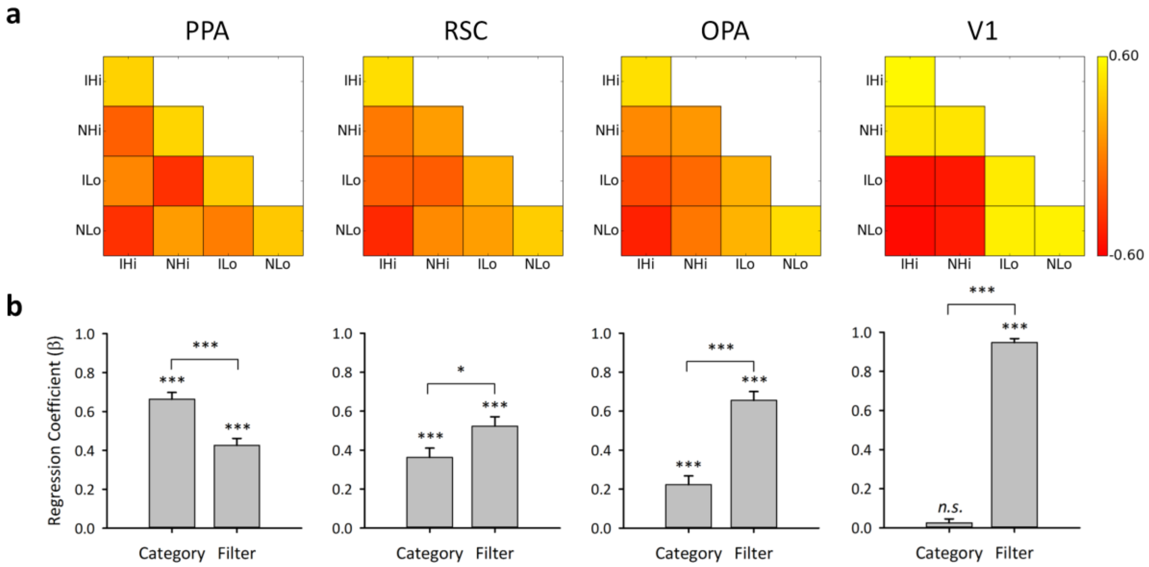


Figure 4.9. Experiment 2: standard scene-selective regions and V1. (a) MVPA correlation matrices. (b) These matrices were compared against binary regressors of category and filter effects using a multiple regression analysis; resulting beta coefficients are shown for each regressor. Error bars represent 1 SEM.
(* p < .05, ** p < .01, *** p < .001).

Table 4.3. Experiment 2: t-statistics and significance of post-hoc pairwise t-tests for standard scene selective regions (PPA, RSC, OPA) and V1.
(* p < .05, ** p < .01, *** p < .001)

	PPA	RSC	OPA	V1
IHi/NHi > IHi/ILo	-3.42**	2.48(ns)	6.12***	18.23***
IHi/NHi > NHi/NLo	-5.55***	1.06(ns)	1.63(ns)	17.33***
ILo/NLo > IHi/ILo	-0.80(ns)	4.10**	7.89***	17.94***
ILo/NLo > NHi/NLo	-2.89*	1.89(ns)	5.49***	16.79***

Previous experiments have suggested a possible division of labour between anterior and posterior regions of the PPA (Aminoff et al., 2007; Epstein, 2008; Arcaro et al., 2009; Baldassano et al., 2013). Accordingly, we re-analysed our data by splitting the PPA region halfway along its posterior-anterior extent and repeating the pattern analyses within each division. Responses in the posterior PPA region were significantly predicted by both the category (β = 0.19, p < .001) and filter regressors (β = 0.63, p < .001), with significantly more variance explained by the filter regressor (t = 5.90, p < .001). Representations in the anterior PPA appeared more similar to the overall PPA region, with responses significantly predicted by both the category (β = 0.75, p < .001) and filter regressors (β = 0.31, p < .001), but with significantly more variance explained by the category regressor (t = 8.51, p < .001). These results are shown in Figure 4.10. Our results therefore show a change in selectivity within the PPA, with a shift from more image-based to more category-based representations along a posterior-to-anterior axis.

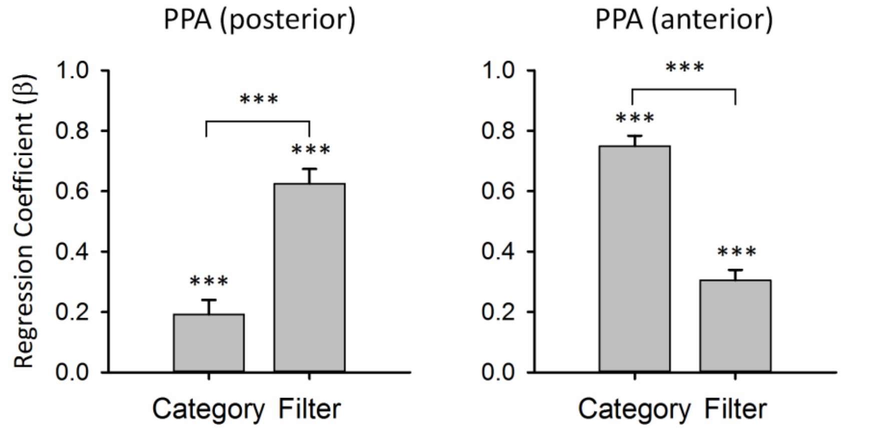


Figure 4.10. Experiment 2: Analysis of anterior and posterior PPA divisions. The PPA region was divided halfway along its posterior-anterior extent, and the pattern analyses and representational analyses repeated for each division separately. The resulting regression coefficients are displayed above. Error bars represent 1 SEM.
(* p < .05, ** p < .01, *** p < .001).

4.4.3 Behavioural Experiment

In order to ensure that the filtering process did not disrupt the ability of participants to perceive the scenes categorically, we conducted an additional behavioural experiment. Participants were presented with the images from the fMRI experiments plus their unfiltered equivalents whilst performing a scene categorisation task. Percentage accuracy scores and median RTs were calculated for each condition within each participant (Table 4.4). Mean accuracy across all conditions was 95.63 ± 1.34% (range 89.17 – 97.92%). Mean RT across all conditions was 598 ± 26ms (range: 566 - 611). These results show that participants were able to categorize all stimulus conditions well above chance levels.

Table 4.4. Behavioural experiment: average accuracy and response times (± 1 SEM).

Category	Filter	Accuracy (% correct)	Response Time (ms)
Indoor	Horizontal-pass	95.83 ± 1.28	597 ± 18
	Vertical-pass	95.42 ± 1.41	611 ± 19
	High-pass	97.08 ± 1.09	569 ± 16
	Low-pass	94.58 ± 1.51	611 ± 18
	Unfiltered	95.42 ± 1.54	566 ± 17
Natural	Horizontal-pass	97.50 ± 1.06	604 ± 23
	Vertical-pass	95.83 ± 1.54	595 ± 21
	High-pass	97.50 ± 1.06	589 ± 16
	Low-pass	89.17 ± 2.10	664 ± 24
	Unfiltered	97.92 ± 0.83	581 ± 20

4.5 Discussion

The aim of this study was to compare the relative effect of low-level image properties and high-level categorical factors on the patterns of fMRI response in scene-selective regions. Participants viewed images from indoor and natural scene categories that were filtered by orientation and spatial frequency. These manipulations had a marked effect on the low level image properties. Nevertheless, a behavioural experiment using stimulus presentation parameters matched to those of the fMRI experiments revealed that these manipulations preserved the ability to accurately categorize the images. We then measured the patterns of response in scene-selective regions. We found that orientation filtering had a significantly smaller effect on patterns of response than category. In contrast, spatial frequency filtering had a significantly greater effect on patterns of response compared to category. These results show that patterns of neural response in scene-selective cortices revealed by fMRI are sensitive to low-level properties of the image, particularly the spatial frequency content.

Previous studies have established that distinct patterns of neural response are elicited by viewing different categories of scene (Walther et al., 2009, 2011). These findings have been taken to suggest a categorical organisation of scene-selective cortices in which response properties are linked to the semantic properties of the image. It has also been shown that the semantic content of scene images can be used to predict neural responses during viewing of natural scenes (Huth et al., 2012; Stansbury et al., 2013) and to reconstruct scene images from neural responses in higher visual areas (Naselaris et al., 2009). However, other studies suggest that categorical factors may not provide a complete account of the organization of scene-selective regions. For instance, reports by both Kravitz et al. (2011) and Park et al. (2011) suggest that responses in PPA are better predicted by image properties (open versus closed) than by the categorical content (indoor versus natural) of scenes. It has also been shown that visual properties can be used to discriminate between different categories of scenes (Torralba & Oliva, 2003). These findings suggest that a fuller understanding of the principles governing organization of ventral visual cortex will hinge on determining the way in which patterns of brain activity reflecting semantic, spatial and functional properties of scenes are derived from their lower level visual properties.

Recently, we showed that the statistical properties of visual images can be used to predict patterns of response in high-level visual cortex (Rice et al., 2014; Watson et al., 2014; Andrews et al., 2015). These results provide an alternative framework for understanding the topographic organization of the ventral visual pathway in which the appearance of category-selective patterns of response may emerge from the combinations of low-level image properties that typically co-occur in different image categories (see also Hanson et al., 2004; Op de Beeck et al., 2008). To directly test the role of image properties, we measured the effect of low-level image manipulations on patterns of response in scene-selective regions. We found a significant effect of spatial frequency filter on patterns of response in scene-selective cortex. For example, indoor low-pass images generated similar patterns of response to natural low-pass images. Similarly, indoor high-pass images generated similar patterns to natural high-pass images. These results show that patterns of response to scenes are sensitive to the low-level properties of the image. Previous univariate fMRI studies have shown that there are biases in the magnitude of the response to different spatial frequencies in scene-selective regions (Rajimehr et al., 2011; Kauffmann et al., 2014). However, changes in the amplitude of response can occur without a change in the pattern of response. Our findings fundamentally extend these earlier studies by showing that the spatial frequency content of the image can also influence the pattern of response in scene-selective regions. This suggests that this property of the image is a key feature underlying the functional organisation of scene-selective regions.

How do we explain the category-specific patterns of response found in scene-selective regions (Walther et al., 2009, 2011)? Rather than reflecting an organization based on categorical properties of the stimulus, we propose that scene-selective regions have a topographic organization that is based on image properties (Andrews et al., 2015). We suggest that the appearance of category selectivity may reflect the characteristic combinations of low-level image properties that co-occur in different types of scenes. Because images from different scene categories have distinct image properties (Watson et al., 2014), images from a particular scene category will activate spatially-selective patterns of response. Although patterns of response in scene-selective regions may be dominated by the features characteristic of specific natural categories, they may remain sensitive to low-level manipulations.

Our findings appear to contrast with a previous study that reported scene category can be decoded from photographs and line drawings of scenes, and that decoding generalises between these visual representations (Walther et al., 2011). As line drawings represent a visually impoverished version of photographic images, it is argued that these results are indicative of image-invariant, categorical representations in scene-selective regions. Our results suggest that such effects could alternatively be understood in terms of the low-level visual properties of images, such as spatial frequency. Line-drawings reduce an image to a subsample of its edge boundaries, and thus represent an extreme high-pass representation of the original image. Consequently, despite being visually impoverished, line drawings will nevertheless maintain similar high spatial frequency content to their original images. Thus, generalisation between each visual representation could reflect sensitivity within the neural patterns to the high spatial frequency content of the image.

Despite showing that manipulations of spatial frequency did affect the patterns of response in scene-selective regions, we also found a smaller but significant effect of scene category across the whole scene-selective ROI. When the scene-selective ROI was subdivided into different sub-divisions (PPA, RSC, TOS/OPA), we found that, although both filter and category influenced the patterns of response, the relative contribution of category and filter varied between regions. For instance, the effect of the spatial frequency filter was greater than that of the category in both the OPA and RSC, while in the PPA the effect of category was greater than filter. This suggests that while all scene-selective regions remain sensitive to the low-level visual properties of scenes, there may be a shift towards a more categorical representation in some regions. Presumably, these differences in selectivity reflect the different computational processes that are thought to occur in different scene-selective regions. For instance, it has been proposed that the PPA and RSC may form distinct but complimentary roles within the scene processing network, with the PPA primarily focussed on representing the spatial components of the immediately visible scene, whilst the RSC is more concerned with representing the scene within the wider spatial environment (Epstein & Higgins, 2007; Epstein et al., 2007a; Epstein, 2008; Park & Chun, 2009). Meanwhile, the more posterior OPA has been proposed to be a lower-level component of a hierarchical scene processing network (Dilks et al., 2013), perhaps analogous to proposed roles for the occipital face area within the face processing network (Haxby et al., 2002). We additionally observed a shift from more image-based to more category-based representations along a posterior-to-anterior axis within the PPA. This suggests an organisation in which representations become less dependent on the individual visual components of images in more anterior regions of parahippocampal cortex, consistent with previous studies suggesting a division of labour along this axis (Epstein, 2008; Baldassano et al., 2013).

In contrast to spatial frequency, we found that manipulating the orientation content of the image had a much smaller effect on the patterns of response across scene-selective cortex. For example, indoor vertical-pass images generated similar patterns of response to indoor horizontal-pass images and natural vertical-pass images generated similar patterns to natural horizontal-pass images. Our results suggest that not all low level properties exert the same degree of influence on large scale patterns of response in scene-selective cortex. This result may seem at odds with a previous study that reported orientation biases in scene-selective regions (Nasr & Tootell, 2012). However, this study differed from our study in two important ways. First, our filters only included the cardinal orientations (horizontal and vertical) and so did not coincide with the cardinal versus oblique orientation bias shown by Nasr and Tootell (2012). Indeed, they did not report any significant differences between cardinal orientations. Second, they used a univariate analysis in which the magnitude of response to cardinal orientations was compared to oblique orientations. In contrast, we investigated the pattern of response across the cortical surface. It is possible to find overall differences in the magnitude of the response between conditions that are not reflected in the pattern of response. So, the finding that the current analyses did not show a significant effect of orientation filtering upon the pattern of response should not be taken as meaning that the regions do not have low-level orientation biases. Rather, it simply means that (horizontal vs. vertical) orientation biases are not found in the pattern of response detected by fMRI.

To understand how the neural representation of scenes changes through the processing hierarchy, we measured the patterns of response in V1. We found that the pattern of response in V1 showed some differences to the patterns found in the scene-selective regions. For instance, while the orientation filters had little effect on the responses in the scene selective regions, a significant effect of both orientation filter and category was found in V1. Furthermore, although a significant effect of both spatial frequency filters and category was observed in scene-selective regions, there was only an effect of spatial frequency filters on the pattern of response in V1. It is important to note, however, that although image filtering techniques do preserve categorical information, they also preserve other visual dimensions that are not influenced by the filtering manipulation. So, the observed effects of the category manipulation may be attributable not only to categorical factors, but also to visual properties that were not affected by the filtering. For example, the effect of category in V1 in Experiment 1 is unlikely to reflect a higher-level representation of scenes in this region, but it is more likely to be driven by differences in the remaining non-orientation-sensitive visual information (such as spatial frequency). Nevertheless, our results indicate a gradual transition in responses to low-level properties such that later processing regions (e.g., PPA) are increasingly sensitive to those features which serve to distinguish behaviourally distinct environments.

In conclusion, in this study we directly determined the effect of low-level image manipulations on the patterns of neural response to different scene categories. We found clear evidence that scene-selective regions were sensitive to the low-level visual content of the image, and that spatial frequency was more influential than orientation content in determining the coarse-scale patterns measured by the MVPA. The sensitivity to image properties shown in this study fundamentally extends previous univariate reports of image biases in the magnitude of response in scene-selective regions. By showing that the pattern of response to scenes can be influenced by the spatial frequency content of the image, our results suggest that this image property is an important organizing factor in the topographic organization of scene-selective regions of the brain.

Chapter 5 – Category-Selective Patterns of Neural Response in Scene-Selective Regions to Intact and Scrambled Images

This chapter is adapted from: Watson, D. M., Hartley, T., & Andrews, T. J. (*in review*). Category-selective patterns of neural response in scene-selective regions to intact and scrambled images. [3]

5.1 Abstract

Neuroimaging studies have found distinct patterns of neural response to different categories of scenes in the human brain. These findings imply that scene category is an important organizing principle in scene-selective regions. However, images from different categories also vary systematically in their lower-level properties. So, it is possible that these patterns of neural response could reflect variance in image properties. To address this question, we used fMRI to measure the patterns of neural response to images of intact scenes and to scenes that had been phase-scrambled at a local or global level. Although both scrambling processes preserved many of the lower-level image properties, categorical perception of the scenes was severely impaired. Nevertheless, we found distinct patterns of response to different scene categories in the parahippocampal place area (PPA) and the occipital place area (OPA) for both intact and scrambled scenes. Moreover, intact and scrambled scenes produced highly similar patterns of response. Our finding that reliable and distinct patterns of response in scene-selective regions are still evident when categorical perception is impaired suggests that the neural representation in these regions may be better explained by the statistical properties of the image.

[3] The author, David Watson, designed the experiment, analysed the results, and wrote the article under the supervision of Dr. Tom Hartley and Prof. Timothy Andrews.

5.2 Introduction

The ability to perceive and recognize the spatial layout of visual scenes is essential for spatial navigation. Neuroimaging studies have identified a number of regions in the human brain that respond selectively to visual scenes (Epstein, 2008). For example, the parahippocampal place area (PPA) is a region of the posterior parahippocampal gyrus that displays preferential activity to images of scenes over and above images of objects and faces (Aguirre et al., 1998; Epstein & Kanwisher, 1998). Other place selective regions include the retrosplenial complex (RSC) located immediately superior to the PPA and the transverse occipital sulcus (TOS) or occipital place area (OPA) on the lateral surface of the occipital lobe (Dilks et al., 2013). Damage to these regions leads to specific impairments in scene perception and spatial navigation (Aguirre & D'Esposito, 1999; Mendez & Cherrier, 2003).

Despite the importance of scene-selective regions for spatial navigation, the functional organisation of these regions remains unclear (Lescroart et al., 2015). Some studies have argued that scene-selective regions represent information about the semantic categories of natural scenes (Walther et al., 2009, 2011; Huth et al., 2012; Stansbury et al., 2013). For example, regions such as the PPA show distinct patterns of response to images of different scene categories (e.g. beaches, forest, buildings). This conclusion has, however, been challenged by other studies that have suggested that the patterns of response in scene-selective regions are better explained by spatial properties of the scene, such as openness (Kravitz et al., 2011; Park et al., 2011) or distance (Amit et al., 2012; Park et al., 2015) rather than by semantic category.

Although concepts such as openness or distance provide a more continuous dimension with which to understand the organization of scene-selective regions, it is not clear whether this can be explained even more simply in terms of low-level image properties that co-vary with these high-level parameters (Oliva & Torralba, 2001). In recent studies, we have shown that variance in the patterns of response to different scene categories can be explained by corresponding variance in the image properties of the scenes (Watson et al., 2014, 2016; Andrews et al., 2015). These findings are consistent with previously reported biases in scene-selective regions for orientation (Nasr

& Tootell, 2012; Nasr et al., 2014), spatial frequency (Rajimehr et al., 2011; Musel et al., 2014) and visual field location (Arcaro et al., 2009; Golomb and Kanwisher, 2012; Levy et al., 2001; Silson et al., 2015) and provide further evidence for the role of image properties in the organization of scene-selective regions. However, images drawn from the same scene category or with the same spatial layout are likely to have similar low-level image properties (Oliva & Torralba, 2001). So, reliable patterns of response are expected under both higher-level and lower-level accounts of scene perception.

The aim of this study was to directly determine whether the patterns of neural response across scene-selective regions can be explained by selectivity to more basic properties of the stimulus. To address this question, we measured the neural response across scene-selective regions to intact images of different scene categories, as well as versions of these images that had been phase-scrambled at a global or local level. Our rationale for using scrambled images is that they have many of the image properties found in intact images, but disrupt perception of categorical and semantic information, thus dissociating high- and low-level information. Our hypothesis was that, if scene-selective regions are selective for the categorical or semantic properties conveyed by the image, there should be no correspondence between patterns of response evoked by intact and scrambled images. Conversely, if patterns of response in scene-selective regions reflect selectivity to more basic dimensions of the stimulus, we would predict a significant correlation between patterns of response to intact and scrambled images.

5.3 Methods

5.3.1 Participants

20 participants (5 males; mean age: 25.85; age range: 19-34) took part in the experiment. All participants were neurologically healthy, right-handed, and had normal or corrected-to-normal vision. Written consent was obtained for all participants and the study was approved by the York Neuroimaging Centre Ethics Committee.

5.3.2 Stimuli

Participants viewed scene images in two independent runs; one to localize the scene-selective regions, the other to experimentally investigate the effects of local and global scrambling manipulations. Images presented in the experiment runs were taken from the LabelMe database (http://cvcl.mit.edu/database.htm; Oliva & Torralba, 2001). Images for the localiser run were taken from the SUN database (http://groups.csail.mit.edu/vision/SUN/; Xiao et al., 2010). Stimuli were presented using PsychoPy (Peirce, 2007, 2009) and were back-projected onto a custom in-bore acrylic screen at a distance of approximately 57 cm from the participant, with all images presented at a resolution of 256x256 pixels subtending approximately 10.7° of visual angle.

The experiment image set comprised 180 greyscale images from 5 scene categories: city, coast, forest, indoor, and mountain (36 images per category). Each image was shown at 3 levels of image scrambling: intact, locally scrambled, and globally scrambled. Globally scrambled images were created by randomising the phase of the 2D frequency components across the whole image whilst keeping the magnitude constant. Locally scrambled images were created by the same process, except that scrambling was applied independently within each of 64 windows of an 8x8 grid across the image. Luminance histograms across all images in all conditions were normalised using the SHINE toolbox (Willenbockel et al., 2010). Examples of the stimuli used in each condition are shown in Figure 5.1.

The localiser images comprised a separate set of 64 scene images plus their phase scrambled counterparts (128 images total), with all images presented in full colour. Fourier-scrambled images were created by randomising the phase of the 2D frequency components in each colour channel of the original image whilst keeping the magnitude constant. Mean luminance was then equated across images.

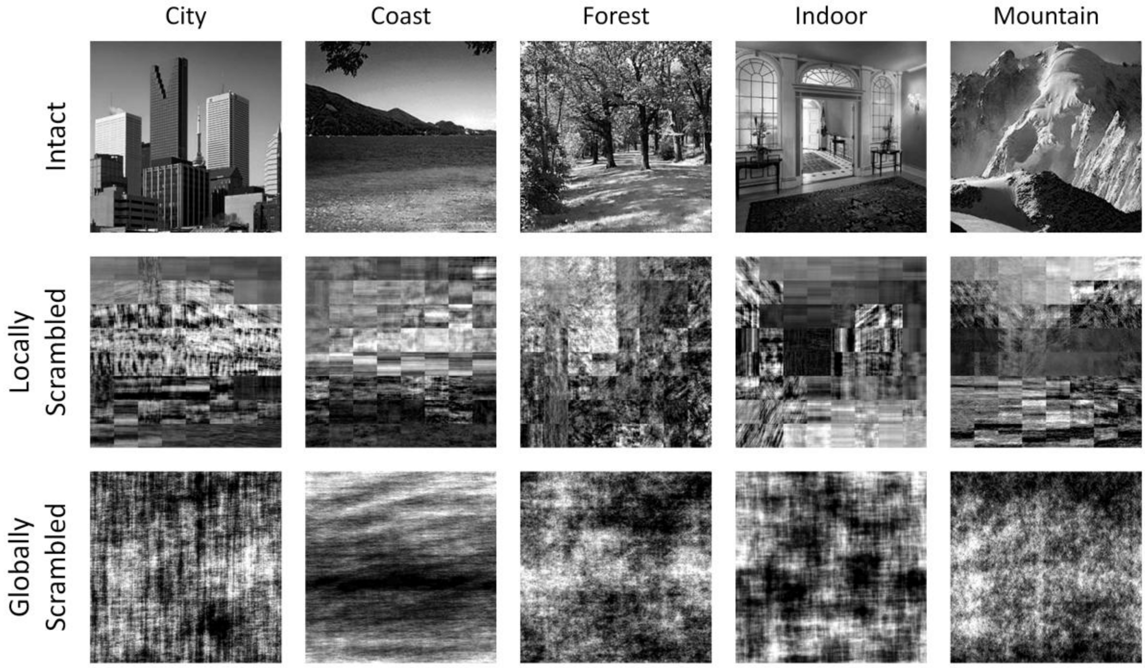


Figure 5.1. Examples of the scene images used in each condition.

5.3.3 fMRI Experimental Design

During the experimental runs participants viewed images from the 5 scene categories. Images from each level of image scrambling were presented across separate experiment runs. For all participants, globally scrambled images were presented in the first run, locally scrambled in the second run, and intact images in the third run. This order was chosen to ensure that responses to scrambled scenes could not be primed by earlier viewing of the intact versions.

In each run, images from each category were presented in a blocked design. There were 6 images in each block. Each image was presented for 750ms followed by a 250ms grey screen that was equal in mean luminance to the scene images. Each stimulus block was separated by a 9s period in which the same grey screen as used in the inter-stimulus interval was presented. Each condition was repeated 6 times (total 30 blocks) in each run. To maintain attention throughout the experimental runs, participants had to detect the presence of a red dot randomly superimposed on one of the images in each block, responding via a button press. Stimuli were presented using PsychoPy (Peirce, 2007, 2009).

To define scene-selective regions, independent data was collected while participants viewed images from 2 stimulus conditions (intact scenes, scrambled scenes). Images from each condition were presented in a blocked fMRI design, with each block comprising 9 images. Each condition was repeated 8 times (16 blocks). In each stimulus block, an image was presented for 750ms followed by a 250ms grey screen. Each stimulus block was separated by a 9s period in which a grey screen was presented. Participants performed a one-back task that involved pressing a button when they detected a repeated image in each block.

5.3.4 Imaging Parameters

All scanning was conducted at the York Neuroimaging Centre (YNiC) using a GE 3 Tesla HDx Excite MRI scanner. Images were acquired with an 8-channel phased-array head coil tuned to 127.72MHz. Data were collected from 38 contigual axial slices in an interleaved order via a gradient-echo EPI sequence (TR = 3s, TE = 32.5ms, FOV = 288x288mm, matrix size = 128x128, voxel dimensions = 2.25x2.25 mm, slice thickness = 3mm with no inter-slice gap, flip angle = 90°, phase-encoding direction = anterior-posterior, pixel bandwidth = 39.06 kHz). In order to aid co-registration to structural images, T1-weighted in-plane FLAIR images were acquired (TR = 2.5s, TE = 9.98ms, FOV = 288x288mm, matrix size = 512x512, voxel dimensions = 0.56x0.56 mm, slice thickness = 3mm, flip angle = 90°). Finally, high-resolution T1-weighted structural images were acquired (TR = 7.96ms, TE = 3.05ms, FOV = 290x290mm, matrix size = 256x256, voxel dimensions = 1.13x1.13 mm, slice thickness = 1mm, flip angle = 20°).

5.3.5 fMRI Analysis

Univariate analyses of the fMRI data were performed with FEAT v5.98 (http://www.fmrib.ox.ac.uk/fsl). In all scans the initial 9s of data were removed to reduce the effects of magnetic stimulation. Motion correction (MCFLIRT, FSL; Jenkinson et al.,

2002) was applied followed by temporal high-pass filtering (Gaussian-weighted least-squares straight line fittings, sigma=15s). Spatial smoothing (Gaussian) was applied at 6mm FWHM to both the localiser and experiment runs, in line with previous studies employing smoothing in conjunction with MVPA (Op de Beeck, 2010; Watson et al., 2014). Parameter estimates were generated for each condition by regressing the hemodynamic response of each voxel against a box-car convolved with a single-gamma HRF. Next, individual participant data were entered into higher-level group analyses using a mixed-effects design (FLAME, FSL). Functional data were first co-registered to an in-plane FLAIR anatomical image then to a high-resolution T1-anatomical image, and finally onto the standard MNI brain (ICBM152).

Scene selective regions of interest (ROIs) were defined from the localiser data of both experiments using the contrast of intact scenes > scrambled scenes. The intact scenes share the same amplitude spectra with their phase scrambled counterparts, thus such a contrast provides a clearer control for low-level visual differences than other commonly used contrasts such as scenes > objects or scenes > faces. For instance, although scenes and objects / faces differ in their category membership, they also differ in a large number of image properties (e.g. spatial frequency, orientation, retinotopic eccentricity, etc.). Given that this experiment aimed to investigate the neural representation of image properties, it was important to use the contrast that provided a stronger control for such visual differences. ROIs were defined for the parahippocampal place area (PPA), retrosplenial complex (RSC), and occipital place area (OPA) that have been reported in previous fMRI studies (Epstein & Kanwisher, 1998; Maguire, 2001; Dilks et al., 2013). Within the MNI-2x2x2mm space, seed points were defined at the peak voxels within the intact>scrambled statistical map for each region (PPA, RSC, OPA) in each hemisphere. For a given seed, a flood fill algorithm was used to identify a cluster of spatially contiguous voxels around that seed which exceeded a given threshold. This threshold was then iteratively adjusted till a cluster size of approximately 500 voxels was achieved (corresponding to a volume of 4000mm^3); actual cluster sizes ranged from 499-502 voxels as an optimal solution to the algorithm was not always achievable. This step ensures that estimates of multi-voxel pattern similarity are not biased by the different sizes of ROIs being compared. Clusters were combined across hemispheres to yield 3

ROIs, each comprising approximately 1000 voxels. These regions are shown in Figure 5.2. MNI co-ordinates of the seeds are given in Table 5.1. These seed points had similar locations to those reported in previous literature (Table A.1).

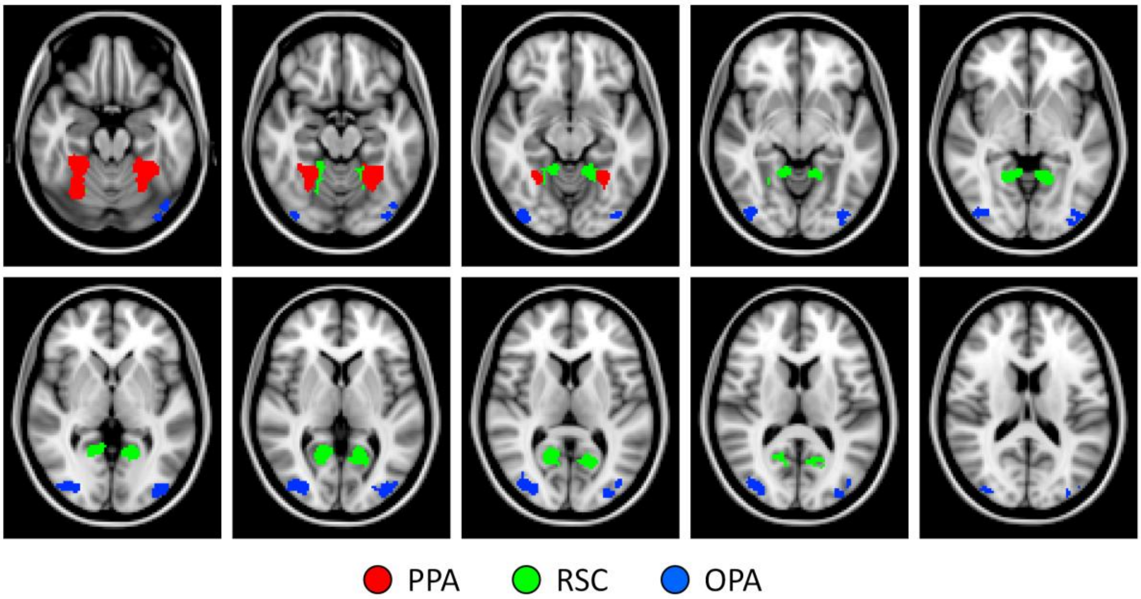


Figure 5.2. Illustration of the masks used for the fMRI analyses. Each mask comprises approximately 500 voxels (4000mm^3) in each hemisphere. Slices of MNI brain span the range from z = -20mm to z = +20mm in 4mm increments.

Table 5.1. Peak MNI mm co-ordinates, voxel counts, and thresholds of standard scene selective clusters (PPA, RSC, OPA).

Region	Hemisphere	*x*	*y*	*z*	Voxel count	Threshold (Z)
PPA	L	-34	-46	-22	500	5.06
	R	26	-50	-18	500	5.59
RSC	L	-18	-52	2	500	4.63
	R	16	-58	6	502	4.79
OPA	L	-36	-90	2	500	5.14
	R	38	-82	4	499	5.03

Next, we measured patterns of response to different stimulus conditions in each ROI. Parameter estimates were generated for each condition in the experimental scans. The reliability of response patterns was tested using a leave-one-participant-out (LOPO) cross-validation paradigm (Shinkareva et al., 2008; Poldrack et al., 2009) in which parameter estimates were determined using a group analysis of all participants except one. This generated parameter estimates for each scene condition in each voxel. This LOPO process was repeated such that every participant was left out of a group analysis once. These data were then submitted to correlation-based pattern analyses (Haxby et al., 2001, 2014) implemented using the PyMVPA toolbox (http://www.pymvpa.org/; Hanke et al., 2009). Parameter estimates were normalised by subtracting the voxel-wise mean response across all experimental conditions (see Haxby et al., 2001). For each iteration of the LOPO cross-validation, the normalized patterns of response to each stimulus condition were correlated between the group and the left-out participant. This allowed us to determine whether there are reliable patterns of response that are consistent across individual participants.

5.3.6 Statistical Analyses

A Fisher's z-transform was applied to the correlation similarity matrices before further statistical analyses. A Bonferroni-Holm correction for multiple comparisons was applied across ROIs.

First, we tested the ability of each region to discriminate the scene categories under each level of image scrambling. For each iteration of the LOPO cross-validation, we calculated an average within-category (on-diagonal) and an average between-category (off-diagonal) value across categories. These values were then entered into a paired-samples t-test. If scene category can be discriminated based on the pattern of activity it elicits, then significantly greater within- than between-category correlations would be expected.

Next, we conducted a series of representational similarity analyses (RSAs; Kriegeskorte et al., 2008) to investigate the effects of different levels of scrambling. Correlation matrices were averaged across iterations of the cross-validation. Representational similarity was assessed by correlating the averaged similarity matrices between the intact and locally scrambled conditions, and between the intact and globally scrambled conditions. If the scrambling does not affect the relative similarity between categories relative to the intact condition, then a significant positive correlation would be expected between the intact and corresponding scrambled matrices.

5.3.7 Behavioural Experiment

We also tested the ability of participants to recognise the scenes under each level of image scrambling. An independent set of 18 participants naive to the purposes of the study were recruited (6 males; mean age: 21.7; age range: 19-39). Written consent was obtained for all participants and the study was approved by the University of York Psychology Department Ethics Committee. Each participant viewed a subset of 1/6th of the image set, comprising 6 images from each category. Subsets were counterbalanced across participants. Participants viewed each image under all three levels of scrambling, and to prevent priming effects participants viewed globally scrambled images first, followed by locally scrambled images, and finally intact images. In each trial participants were shown an image for as long as they wished and were required to describe the type of scene they thought was shown, typing responses into a text box below the image. Participants were free to provide any description they wanted, and were also informed that they did not have to give a response if they could not reasonably see what type of scene was depicted. Accuracy was coded manually, and a correct response was defined as any which could reasonably be seen to accurately describe the corresponding intact scene. Accuracies were converted to proportions and an arcsine square-root transform was applied prior to further statistical tests. If participants did provide a description, they were next prompted to provide a confidence rating of their decision on a 7 point scale (not at all confident - very confident). Participants were not provided with any information about the scene categories prior to the experiment; this is important as participants in the fMRI experiment were not given any prior knowledge of the scene categories either.

5.4 Results

5.4.1 Behavioural Experiment

We first tested the effects of the different levels of scrambling on participants' ability to recognise the scenes. Mean accuracy for each condition is shown in Figure 5.3a. As expected, accuracy was higher for intact (mean = 100 ± 0%) compared to locally scrambled (mean = 31.48 ± 3.37%) and globally scrambled images (mean = 4.63 ± 0.86%). A one-way repeated measures ANOVA revealed a significant main effect of scrambling (F(2,34) = 811.17, p < .001, generalized-η^2 = .97). A series of post-hoc t-tests revealed significantly higher accuracies for intact compared to locally scrambled scenes, intact compared to globally scrambled scenes, and locally scrambled compared to globally scrambled scenes (all p < .001). Participants also provided confidence ratings of their descriptions on a scale of 1 (not at all confident) to 7 (very confident). Median ratings for each condition were calculated for each participant and are shown in Figure 5.3b. Similar to the accuracy data, confidence ratings were higher for intact (median = 7, IQR = 6 - 7) compared to locally scrambled (median = 3, IQR = 2 - 4) and globally scrambled images (median = 2, IQR = 2 - 2.5). A Friedman's ANOVA revealed a significant main effect of scrambling ($\chi^2(2)$ = 32.62, p < .001). A series of post-hoc Wilcoxon signed-rank tests revealed significantly higher confidence ratings for intact than locally scrambled scenes (p < .001), intact than globally scrambled scenes (p < .001), and locally scrambled than globally scrambled scenes (p = .002). Thus both types of scrambling significantly impaired participants' recognition and confidence on a scene recognition test.

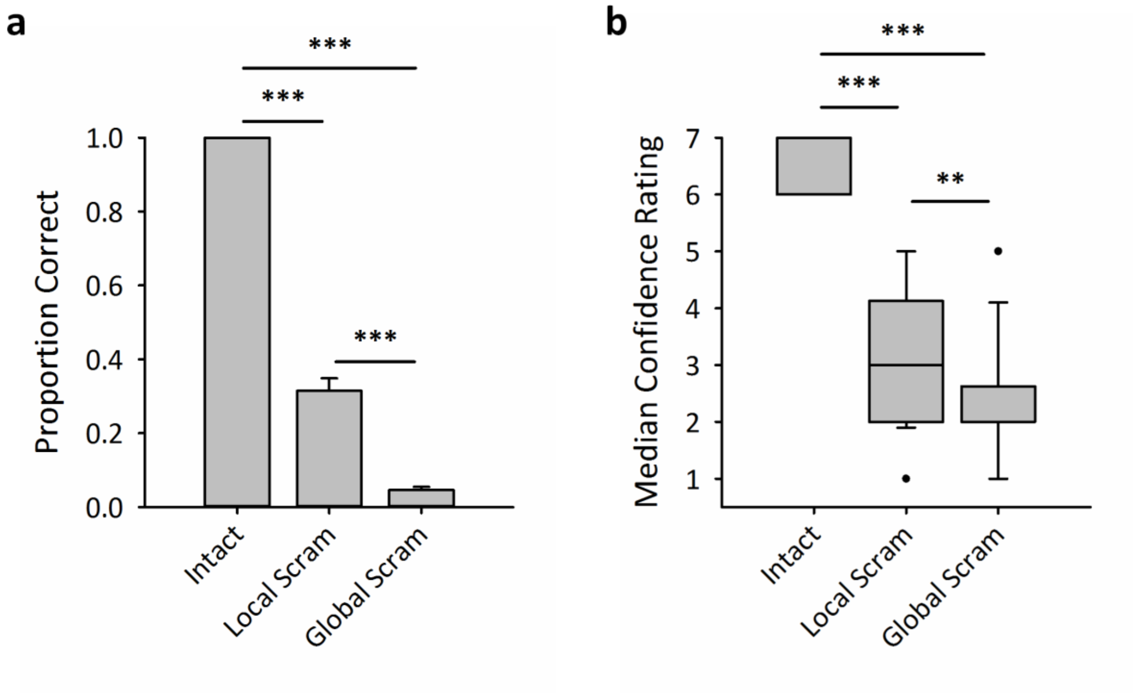


Figure 5.3. Results of the behavioural experiment. (a) Mean scene identification accuracies for each level of scrambling. Error bars represent 1 SEM. (b) Box-plots of median confidence ratings for each level of scrambling.

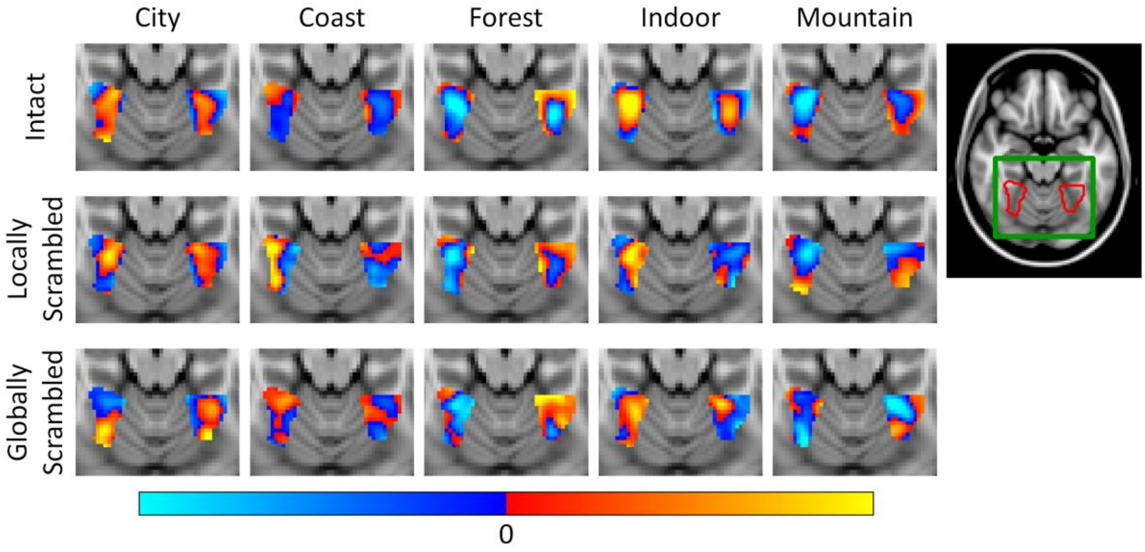


Figure 5.4. Group patterns of response for each condition, restricted to PPA region. Responses within each level of scrambling are normalized by subtracting a voxel-wise mean across all categories, such that red and blue colours indicate values above and below the mean respectively.

5.4.2 fMRI experiment

Next, we used fMRI to measure the patterns of neural response to each of the conditions. The group normalised responses within the PPA region are shown in Figure 5.4 (red and blue colours indicate responses above and below the mean respectively). Responses within the RSC and OPA regions are shown in Figures A.13 and A.14. Correlation-based MVPA (Haxby et al., 2001) using a leave-one-participant-out (LOPO) cross-validation scheme was then used to assess the reliability of these responses. Average correlation similarity matrices for each of the ROIs and each of the scrambling types are shown in Figure 5.5, with symmetrically opposite points averaged across the diagonal to aid visualisation.

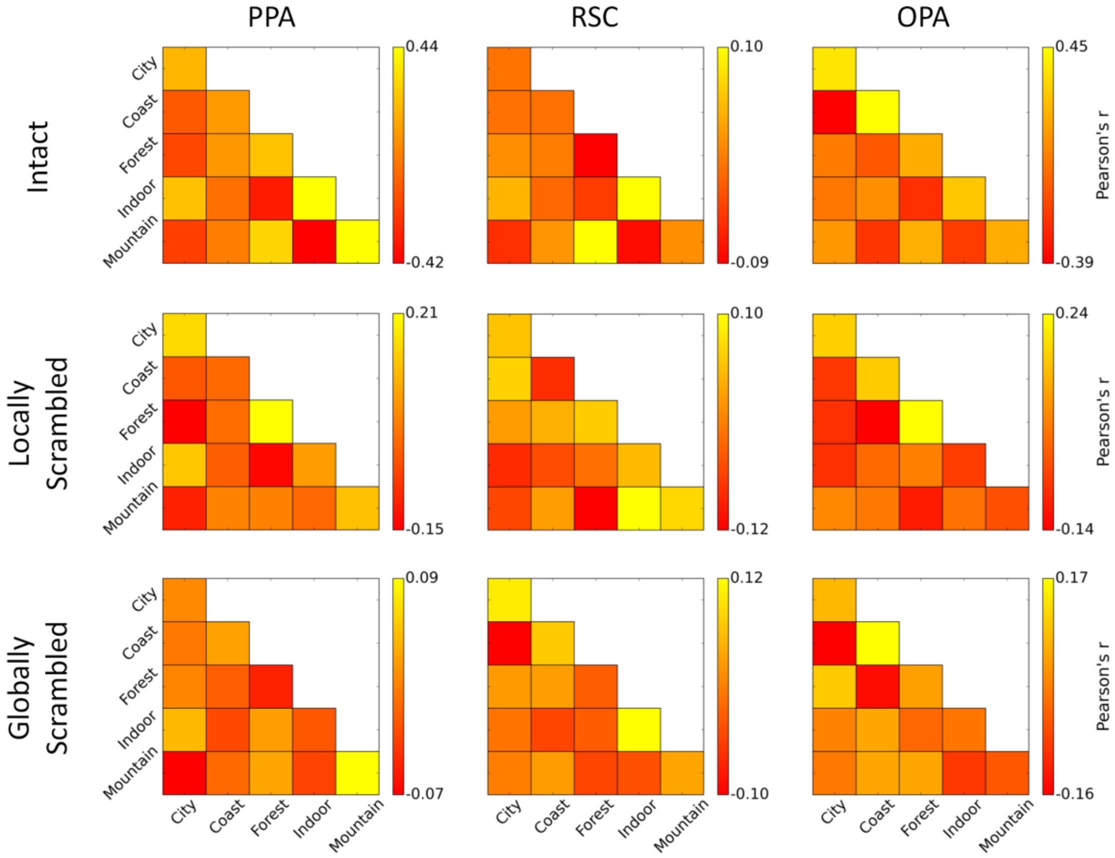


Figure 5.5. MVPA results: correlation similarity matrices for each level of scrambling in each region of interest. To aid visualisation, symmetrically opposite points across the diagonal have been averaged and displayed within the lower-triangle portion of the matrix only.

We first assessed the ability of the MVPA to discriminate the scene categories under each of the levels of scrambling. We calculated within- and between-category correlation values averaged across categories for each scrambling type and ROI. These values are shown in Figure 5.6. Paired-samples t-tests were then used to test for differences between within- and between-category correlations; if categories can be discriminated based on patterns of brain activity, then significantly greater within- than between-category correlations would be expected. For the intact scenes, significantly greater within- than between-category correlations were observed in the PPA (t(19) = 10.90, *p* < .001, Cohen's d = 2.44) and OPA (t(19) = 9.89, *p* < .001, Cohen's d = 2.21), but not in the RSC (t(19) = 0.17, *p* > .999, Cohen's d = 0.04). In the locally scrambled condition, significantly greater within- than between-category correlations were found in the PPA (t(19) = 5.54, *p* < .001, Cohen's d = 1.24) and OPA (t(19) = 4.57, *p* = .001, Cohen's d = 1.02), but not in the RSC (t(19) = 1.43, *p* = .498, Cohen's d = 0.32). For the globally scrambled scenes, no significant differences were seen for any ROI (PPA: t(19) = 0.43, *p* > .999, Cohen's d = 0.10; RSC: t(19) = 2.20, *p* = .200, Cohen's d = 0.49; OPA: t(19) = 2.14, *p* = .200, Cohen's d = 0.48).

We next conducted a series of representational similarity analyses (RSAs; Kriegeskorte et al., 2008) to test to what extent the two types of scrambling influence the representational structure of the responses relative to those of the intact scenes. The group average matrices (each comprising 25 elements) were correlated between intact and locally scrambled conditions, and intact and globally scrambled conditions. If the scrambling does not disrupt the representational space, a significant positive correlation would be expected with the intact scenes matrix. A significant positive correlation was observed between intact and locally-scrambled scenes in the PPA (*r* = .72, *p* < .001), but not in the RSC (*r* = -0.43, *p* = .132) or OPA (*r* = .31, *p* = .250). A significant positive correlation was observed between intact and globally scrambled conditions in the OPA (*r* = .66, *p* = .002), but not the PPA (*r* = .40, *p* = .151) or RSC (*r* = .21, *p* = .325). These results are illustrated in Figure 5.7.

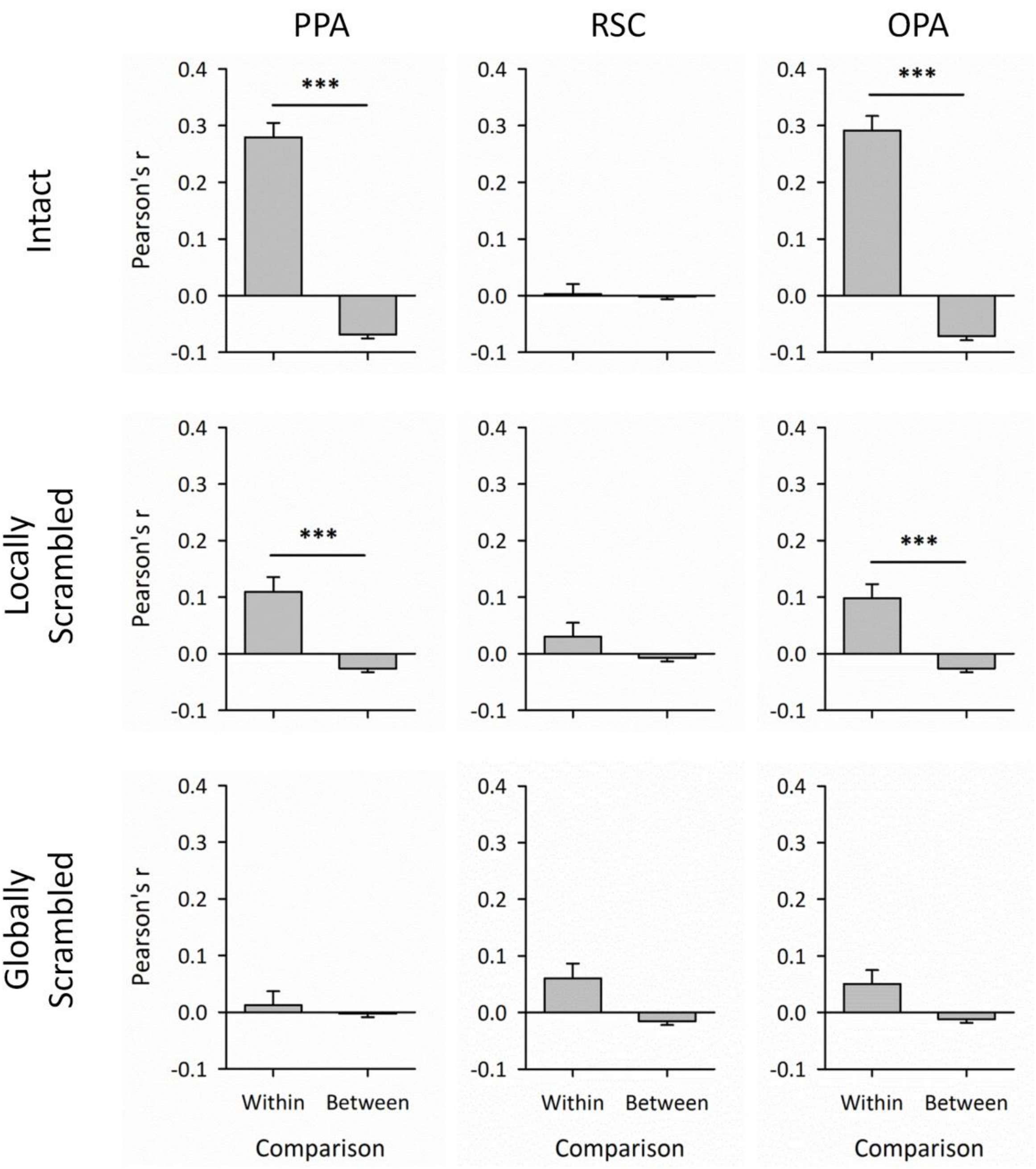


Figure 5.6. Decoding of categories from MVPA. Average within-category (on-diagonal) and between-category (off-diagonal) values are calculated from the MVPA correlation matrices. Significantly greater within- than between-category correlations indicate categories can be successfully decoded (* p < .05, ** p < .01, *** p < .001).

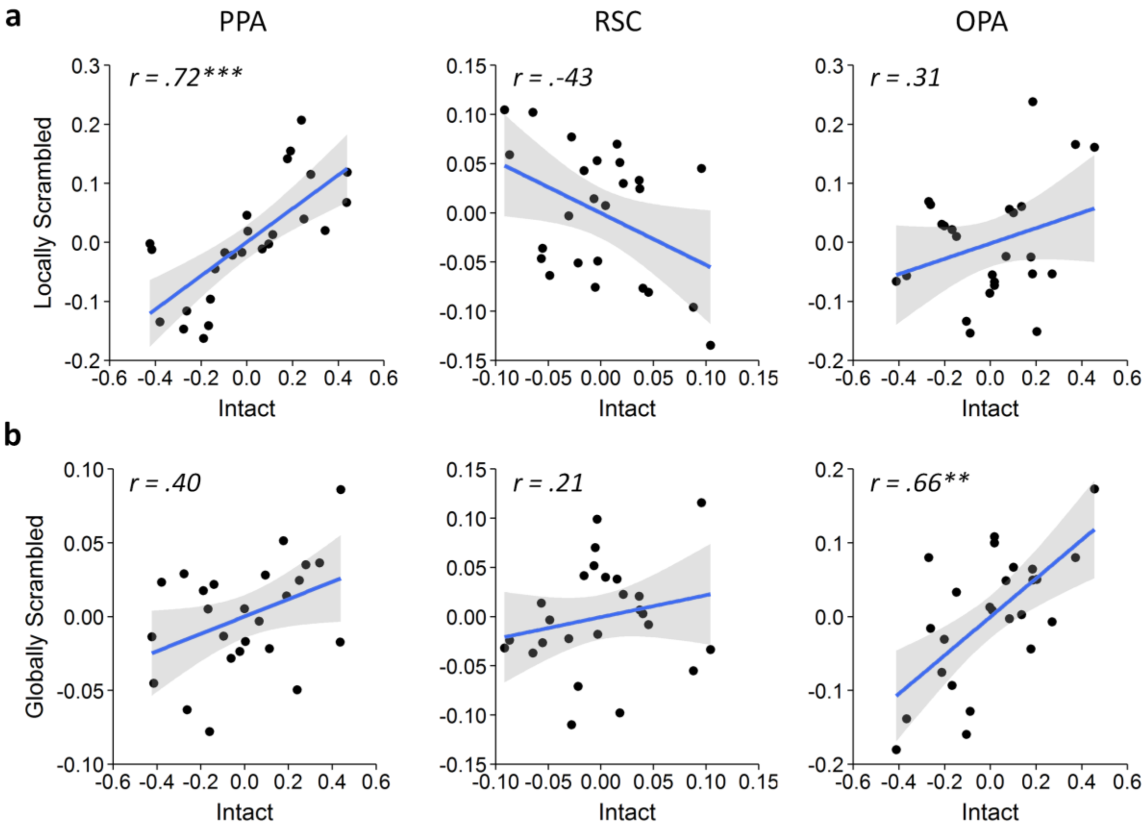


Figure 5.7. Representational similarity analyses. Group average MVPA correlation matrices (Figure 5.5) are correlated between intact and locally-scrambled conditions, and between intact and globally-scrambled conditions. Shaded regions represent 95% confidence intervals (* p < .05, ** p < .01, *** p < .001).

Correlation values between intact and scrambled matrices could reflect the distinction between higher within-category (on-diagonal) compared to between-category (off-diagonal) elements, regardless of the underlying representational structure. To address this issue, we repeated our analyses, but restricted the analysis to only the off-diagonal elements of the matrices. A similar pattern of correlations was found both when comparing intact and locally scrambled conditions (PPA: r = .66, p = .009; RSC: r = -0.56, p = .044; OPA: r = -.15, p > .999), and intact and globally scrambled conditions (PPA: r = .43, p = .160; RSC: r = .02, p > .999; OPA: r = .62, p = .019).

5.5 Discussion

The aim of the present study was to directly determine whether category-selective patterns of response in scene-selective regions were better explained by scene category or by more basic dimensions of the stimulus. To address this issue, we compared patterns of response to intact and scrambled images. Our hypothesis was that if category-selective patterns of response reflect the categorical or semantic content of the images, there should be little similarity between the patterns of response elicited by intact and scrambled images. On the other hand, if category-specific patterns are based on more basic image properties, similar patterns should be elicited by both intact and scrambled images. Image scrambling significantly impaired the ability to categorize scenes. However, we found distinct and reliable category-selective patterns of response for both the intact and scrambled image conditions in the PPA and OPA regions, but not the RSC. Moreover, the patterns of response elicited by intact scenes were similar to the patterns of response to scrambled scenes. This was most evident between intact and locally scrambled scenes in the PPA, and between intact and globally scrambled scenes in the OPA.

Previous studies have identified distinct patterns of neural response to different categories of scene in scene selective regions (Walther et al., 2009, 2011). These results have been taken to suggest that such regions may play a role in the categorisation of scenes (Walther et al., 2009). Our results show that categorical patterns of response in scene-selective regions are still evident to images with significantly reduced categorical information. This suggests that the topographic organization in regions such as the PPA is based on more fundamental properties of the image. These findings are consistent with recent studies in which we have shown that basic image properties of different scene categories can predict patterns of response in scene-selective regions (Watson et al., 2014). However, because images drawn from the same category are likely to have similar lower-level properties, it was unclear from this previous work whether patterns are determined primarily by membership of a common category or by the shared lower-level image statistics characteristic of that category (Lescroart et al., 2015). The results from the current study provide more direct evidence that lower-level properties of the image can account for patterns of response in scene-selective regions.

To evaluate the importance of spatial properties in the neural representation of scenes, we compared scrambling across the full global extent of the image, or independently within local windows of the image. The local scrambling thus preserves the spatial organisation of the original image more than the global scrambling. In the PPA, we found that responses could be discriminated for locally, but not globally scrambled scenes. Furthermore, a representational similarity analysis showed that local scrambling, rather than global scrambling best preserved the relative similarity in response relative to the intact scenes. This would suggest that the PPA is sensitive to the local spatial organisation of the image, such that responses are more severely disrupted by globally scrambling the image. Such a conclusion would be consistent with previous studies demonstrating sensitivity of the PPA to the spatial structure of scenes (Epstein et al., 2006; Kravitz et al., 2011; Park et al., 2011), and displaying visual field biases (Arcaro et al., 2009; Cichy et al., 2013; Silson et al., 2015). Indeed, it has been proposed that the PPA may support extraction of local spatial geometries of the scene (Epstein et al., 2007a; Epstein, 2008), for which local visual features may be important.

We found that category responses in the OPA could also be discriminated for intact scenes and locally scrambled scenes, but not globally scrambled scenes. However, in contrast to the PPA the representational similarity analysis showed that the representational structure of the intact scenes was better maintained by the global than the local scrambling. Although the OPA has been causally implicated in the perception of scenes (Dilks et al., 2013; Ganaden et al., 2013), its precise functional properties are less well established than other scene regions. It should be noted that the local scrambling process does also introduce some disruption to the global features of the image that the global scrambling does not; for instance high spatial frequency artifacts are introduced at the edges between windows, and the phase coherence of components spanning multiple windows is not maintained. Thus responses in a region sensitive to the global statistics of the image may still be disrupted by the local scrambling. Nevertheless, our results do support the idea that OPA responses demonstrate sensitivity to visual features of scenes even when scene perception is disrupted. Furthermore, they suggest a possible functional distinction between PPA and OPA, with the PPA more clearly tuned to the local visual features than the OPA.

In contrast to the PPA and OPA, responses in RSC failed to discriminate the scene categories in any of the conditions. The representational similarity analyses showed that neither local nor global scrambling maintained the representational structure relative to the intact scenes. It has been proposed that the RSC may play a role representing the scene as part of the wider spatial environment (Epstein et al., 2007a; Epstein, 2008) playing a crucial role in spatial memory, navigation and imagery – for example, translating between ego- and allocentric spatial representations (Byrne et al., 2007; Vann et al., 2009). Such processes may be expected to be less dependent on the immediate visual features of the scene, but at the same time are likely to be more severely disrupted by impaired perception of the scene. Thus, it might be expected that scrambling the scene would disrupt response patterns relative to those of intact scenes.

In conclusion, our results demonstrate distinct responses to different categories of scenes even when the perception of scene category is severely impaired. These results suggest that semantic category may not be a dominant organizing principle in scene-selective regions. Rather, they suggest that the neural representations in these regions may be better explained by the statistical properties of the image.

Chapter 6 – A Data Driven Analysis Reveals the Importance of Image Properties in the Neural Representation of Scenes

This chapter is adapted from: Watson, D. M., Andrews, T. J., Hartley, T. (*in review*). A data driven analysis reveals the importance of image properties in the neural representation of scenes. [4]

6.1 Abstract

The neural representation of scenes in human visual cortex has been linked to processing of semantic and categorical properties (e.g. categorization of indoor versus outdoor scenes). However, it is not clear whether patterns of neural response in these regions reflect more fundamental visual principles like those that govern the organization of early visual cortex. One problem is that existing studies have involved comparisons between stimulus categories chosen by the experimenter, potentially obscuring the contribution of more basic visual features. Here, we used a data-driven analysis to select clusters of scenes based solely on their image properties. Although these visually-defined clusters did not correspond to conventional scene categories, we found they elicited distinct and reliable patterns of neural response, and that the relative similarity of the response patterns to different clusters could be predicted by the low-level properties of the images. Local semantic properties of the images failed to explain any additional variance in the neural responses of scene-selective regions beyond that explained by the image properties. However, we did find that participants' behavioural classification of the scenes was better predicted by local semantic properties than by image properties. These results suggest that image properties play an important part in governing patterns of response to scenes in high-level visual cortex and suggest that these patterns are at least

[4] The author, David Watson, designed the experiment, analysed the results, and wrote the article under the supervision of Prof. Timothy Andrews and Dr. Tom Hartley.

partially dissociated from behavioural responses which are better explained in terms of local semantic content.

6.2 Introduction

Human observers are reliably able to perceive and categorize scenes (for example indoor, outdoor) based on their spatial organisation and semantic content. These processes are thought to rely upon a network of regions in the human brain that have been shown to respond preferentially to images of spatial scenes (Aguirre & D'Esposito, 1997; Epstein & Kanwisher, 1998; Dilks et al., 2013). While studies using univariate fMRI analyses have reported comparable overall *magnitudes* of response within these regions to different scene categories (Epstein & Kanwisher, 1998), more recent reports employing multivariate techniques have identified distinct *patterns* of response to different types of scene (Walther et al., 2009, 2011; Marchette et al., 2015) suggesting a finer-grained organisation within scene-selective regions.

Although it is clear that participants can perceive and distinguish scene categories, and that patterns of neural response reflect categorical distinctions, it is by no means obvious that neural responses are systematically organised by semantic category, or that the perception of categories and categorical behavioural responses are causally linked to such patterns (Lescroart et al., 2015). Indeed, recent studies have suggested that patterns of response in scene-selective regions may be better explained in terms of visual properties of scenes, related to spatial characteristics, such as openness (Kravitz et al., 2011; Park et al., 2011) or distance (Amit et al., 2012; Park et al., 2015) rather than by semantic category. These studies are not explicit about how the image properties of the scene are linked to the patterns of neural response, but work in computer vision indicates that semantically-distinct scene categories can be identified on the basis of their low-level image statistics. For example, the visual properties of the image can be used to accurately classify different scene categories and derive spatial properties such as openness (Torralba & Oliva, 2003). Recently, we showed that the same visual properties also predict the topographic pattern of response in scene-selective regions (Watson et al., 2014). Furthermore, we showed that direct manipulations of low-level features have a marked effect on the pattern of response in scene-selective regions (Watson et al., 2016). These findings suggest that patterns of response in scene-selective regions may be determined by more basic dimensions of the stimulus, perhaps similar to those that govern the functional topography of early visual cortex, rather than high-level semantic or categorical properties.

To understand how the perception of scene categories might emerge from more basic visual characteristics of images, we set out to investigate their contribution to the patterns of response in scene selective regions. A fundamental problem in almost all univariate and multivariate studies to date is that they have relied on experimental designs employing experimenter-defined stimulus categories. This makes it difficult to separate the effects of the arbitrary and subjective manipulation of semantic category from those driven by correlated image statistics. However, if the underlying organisational principles governing patterns of neural response in scene-selective cortex draw on such low-level properties, the coupling of neural responses and visual descriptors should persist even when the stimuli are selected solely on the basis of their visual characteristics.

In this study, we directly compared the relative contribution of image properties and semantic features to the organization of scene-selective regions using a data-driven approach in which images are selected based only on their visual content. We used a measure of visual properties (GIST; Oliva and Torralba, 2001) in conjunction with an unsupervised learning algorithm to identify clusters of scenes according to their visual content. If scene-selective areas are sensitive to the visual content of scenes independent of semantic properties, we would expect to find: 1) distinct patterns of response to each scene cluster, 2) the similarity of neural responses to different scene clusters is well explained by the similarity of the corresponding visual descriptors, and 3) semantic properties of the different scene clusters do not explain any additional variance in the neural responses beyond that explained by the visual descriptors.

6.3 Methods

6.3.1 Participants

20 participants (5 males; mean age: 25.8; age range: 19-34) took part in the experiment. All participants were neurologically healthy, right-handed, and had normal or corrected-to-normal vision. Written consent was obtained for all participants and the study was approved by the York Neuroimaging Centre Ethics Committee.

6.3.2 Data-Driven Image Selection

The experimental stimulus set was generated by an entirely data-driven approach. In order to reflect the high variability of real world scenes we selected images from the SUN397 database (Xiao et al., 2010) as this offers a large number (over 100,100 images) and diverse range of scenes. Image properties were measured with the GIST descriptor (Oliva & Torralba, 2001) as this has previously been shown to provide a good model of neural responses in scene selective regions (Rice et al., 2014; Watson et al., 2014; Andrews et al., 2015). The GIST descriptor uses a vector of 512 values to represent an image in terms of the spatial frequencies and orientations present at different spatial locations across the image (Figure 6.1a).

Images were first cropped and resized to the resolution that they would be presented at in the experiment (256x256 pixels), and converted to grayscale. A GIST descriptor was then calculated for every image in the SUN database. GIST vectors were next normalised by first scaling each component of the vectors to sum to 1 across images, and second by scaling each vector to have a magnitude of 1. Each image is thus represented as a point in a 512-dimensional feature space by its normalised GIST descriptor. Attempting to apply clustering algorithms in such a high-dimensional space can be problematic, so we first reduced the dimensionality using principal components analysis (PCA). The first 20 principal components were selected; these explained 70.35% of the variance of the original components. We applied a k-Means clustering algorithm (k = 10; Euclidean distance metric) to identify 10 distinct clusters of samples within this space, such that samples within a cluster are defined by having similar image properties to one another. Finally, we selected the 24 points nearest the centroid of each cluster as measured by Euclidean distance. This process is illustrated in Figure 6.1b. The GIST descriptor is not sensitive to colour, so images were presented in greyscale. Mean luminance and visual contrast were equated across images. Examples of images from each cluster are shown in Figure 6.2.

To help visualise the structure of the points within the feature space, we computed a correlation based similarity matrix using a leave-one-image-out cross-validation procedure. For each cluster, the principal component vectors were averaged across all but one of the images, and the average and left-out vectors correlated within and between clusters. This process was then repeated so that every image was left out once. Figure 6.1c shows the correlations matrix averaged across the cross-validation iterations. We also used multi-dimensional scaling (MDS) to provide a 2D visualisation approximating the distribution of samples within the feature space (Figure 6.1d). PCA, k-Means, and MDS algorithms were all implemented using the Python Scikit-learn toolbox (Pedregosa et al., 2011).

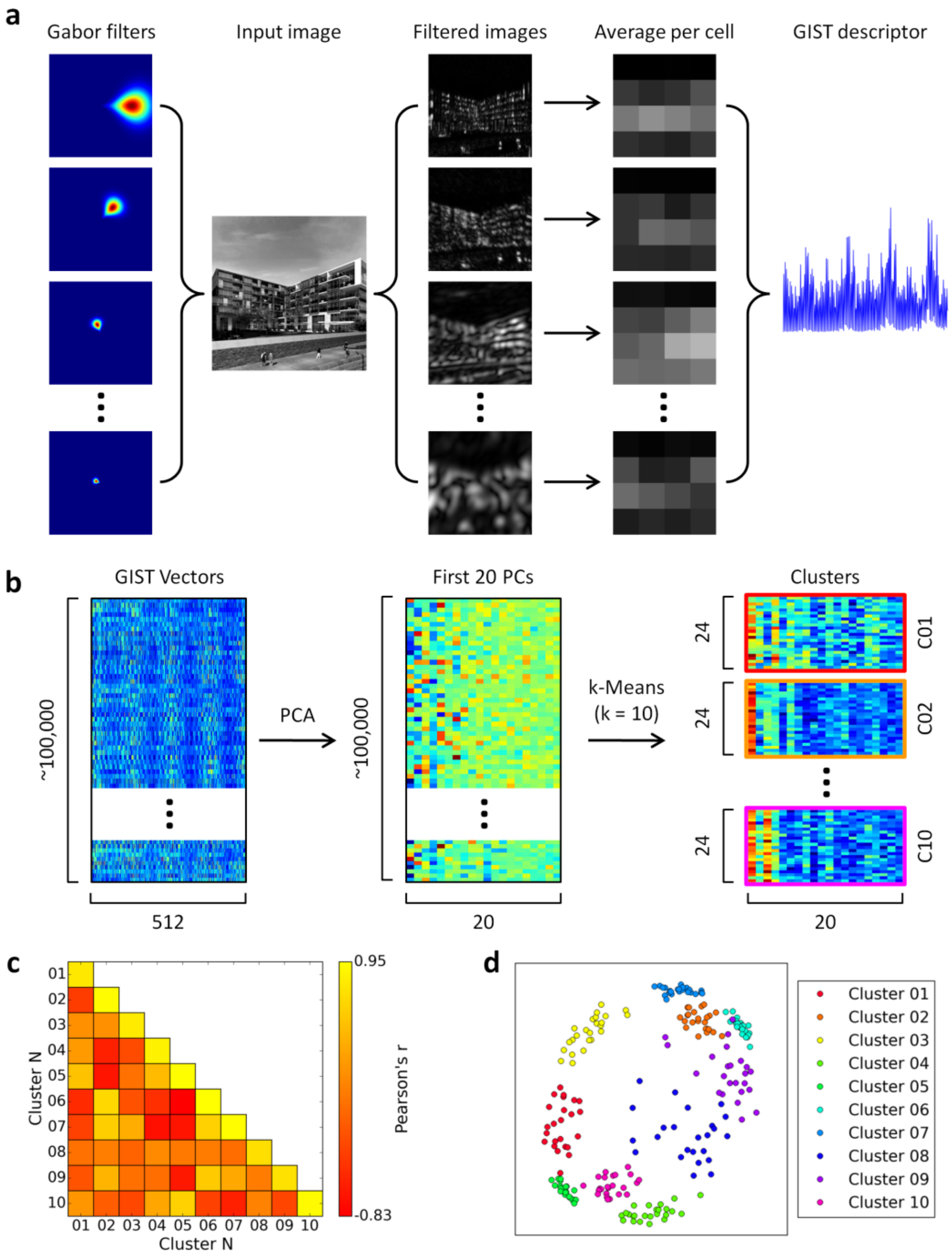


Figure 6.1. GIST clustering process. (a) The GIST descriptor (Oliva & Torralba, 2001) comprises a vector of 512 values that represent the image in terms of the spatial frequencies and orientations present within each of 16 spatial locations across the image. (b) GIST descriptor vectors were calculated for every image in the SUN database. PCA was used to reduce dimensionality down to the first 20 components, and a k-Means clustering algorithm then used to select 10 clusters of scenes. Finally, the 24 images nearest the centroid of each cluster were selected to form the final stimulus set. The structure of the feature space is illustrated by the correlations similarity matrix (c) and multi-dimensional scaling plots (d).

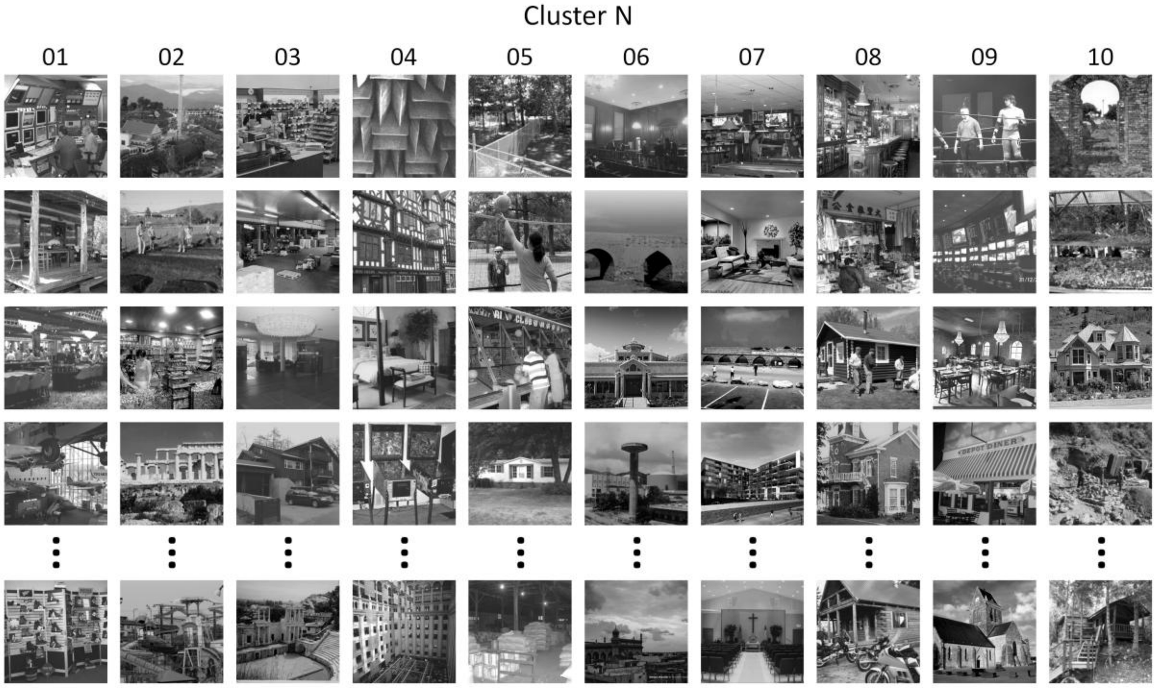


Figure 6.2. Examples of the stimuli from each of the scene clusters.

6.3.3 fMRI Experimental Design

Visual stimuli were back-projected onto a custom in-bore acrylic screen at a distance of approximately 57 cm from the participant, with all images presented at a resolution of 256x256 pixels subtending approximately 10.7° of visual angle. Images presented in both the experiment and localiser runs were taken from the SUN database (http://groups.csail.mit.edu/vision/SUN/; Xiao et al., 2010). Stimuli were presented using PsychoPy (Peirce, 2007, 2009).

During the experimental scan participants viewed images from the 10 scene clusters. Images from each condition were presented in a blocked fMRI design, with each block comprising 6 images. Each image was presented for 750ms followed by a 250ms grey screen that was equal in mean luminance to the scene images. Each stimulus block was separated by a 9s period in which the same grey screen as used in the inter-stimulus interval was presented. Each condition was repeated 4 times giving a total of 40 blocks. To maintain attention throughout the scan participants performed a passive task detecting the presence of a red dot randomly superimposed on one of the images in each block, responding via a button press.

An independent localiser scan was used to define scene-selective regions. During the localiser scan, participants viewed images from 2 stimulus conditions: 1) intact scene images and 2) phase scrambled versions of the same images in condition 1. Images from each condition were presented in a blocked fMRI design, with each block comprising 9 images. Each block was separated by a 9s period in which the same grey screen was presented. Each condition was repeated 8 times giving a total of 16 blocks. To maintain attention participants performed a one-back task detecting the presentation of a repeated image in each block, responding via a button press.

6.3.4 Imaging Parameters

All scanning was conducted at the York Neuroimaging Centre (YNiC) using a GE 3 Tesla HDx Excite MRI scanner. Images were acquired with an 8-channel phased-array head coil tuned to 127.72 MHz. Data were collected from 38 contigual axial slices in an interleaved order via a gradient-echo EPI sequence (TR = 3s, TE = 32.5ms, FOV = 288x288mm, matrix size = 128x128, voxel dimensions = 2.25x2.25 mm, slice thickness = 3mm with no inter-slice gap, flip angle = 90°, phase-encoding direction = anterior-posterior, pixel bandwidth = 39.06 kHz). In order to aid co-registration to structural images, T1-weighted in-plane FLAIR images were acquired (TR = 2.5s, TE = 9.98ms, FOV = 288x288mm, matrix size = 512x512, voxel dimensions = 0.56x0.56 mm, slice thickness = 3mm, flip angle = 90°). Finally, high-resolution T1-weighted structural images were acquired (TR = 7.96ms, TE = 3.05ms, FOV = 290x290mm, matrix size = 256x256, voxel dimensions = 1.13x1.13 mm, slice thickness = 1mm, flip angle = 20°).

6.3.5 fMRI Analysis

Univariate analyses of the fMRI data were performed with FEAT v5.98 (http://www.fmrib.ox.ac.uk/fsl). In all scans the initial 9s of data were removed to reduce the effects of magnetic stimulation. Motion correction (MCFLIRT, FSL; Jenkinson et al., 2002) was applied followed by temporal high-pass filtering (Gaussian-weighted least-squared straight line fittings, sigma=15s). Spatial smoothing (Gaussian) was applied at 6mm FWHM to both the localiser and experiment runs, in line with previous studies employing smoothing in conjunction with MVPA (Op de Beeck, 2010; Watson et al., 2014). Parameter estimates were generated for each condition by regressing the hemodynamic response of each voxel against a box-car regressor convolved with a single-gamma HRF. Head motion parameters were also included as confound regressors. Next, individual participant data were entered into higher-level group analyses using a mixed-effects design (FLAME, FSL). Functional data were first co-registered to an in-plane FLAIR anatomical image then to a high-resolution T1-anatomical image, and finally onto the standard MNI brain (ICBM152).

Scene-selective regions of interest (ROIs) were defined from the localiser data of both experiments using the contrast of intact scenes > scrambled scenes. Because intact scenes share the same amplitude spectra with their phase scrambled counterparts, this contrast provides a clearer control for low-level visual differences than other commonly used contrasts such as scenes > objects or scenes > faces. For instance, although scenes and objects / faces differ in their category membership, they also differ in a large number of image properties (e.g. spatial frequency, orientation, retinotopic eccentricity, etc.). Given that this experiment aimed to investigate the neural representation of image properties, it was important to use the contrast that provided a stronger control for such visual differences. ROIs were defined for the parahippocampal place area (PPA), retrosplenial complex (RSC), occipital place area (OPA) that have been reported in previous fMRI studies (Epstein & Kanwisher, 1998; Maguire, 2001; Dilks et al., 2013). Within the MNI-2x2x2mm space, seed points were defined at the peak voxels within the intact>scrambled statistical map for each region (PPA, RSC, OPA) in each hemisphere. For a given seed, a flood fill algorithm was used to identify a cluster of spatially contiguous voxels around that seed which exceeded a given threshold. This threshold was then iteratively adjusted till a cluster size of approximately 500 voxels was achieved (corresponding to a volume of 4000mm^3); actual cluster sizes ranged from 499-502 voxels as an optimal solution to the algorithm was not always achievable. This step ensures that estimates of multi-voxel pattern similarity are not biased by the different sizes of ROIs being compared. Clusters were combined across hemispheres to yield 3 ROIs, each comprising approximately 1000 voxels. These regions are shown in Figure 6.3. MNI co-ordinates of the seeds are given in Table 6.1. These seed points had similar locations to those reported in previous literature (see Table A.1).

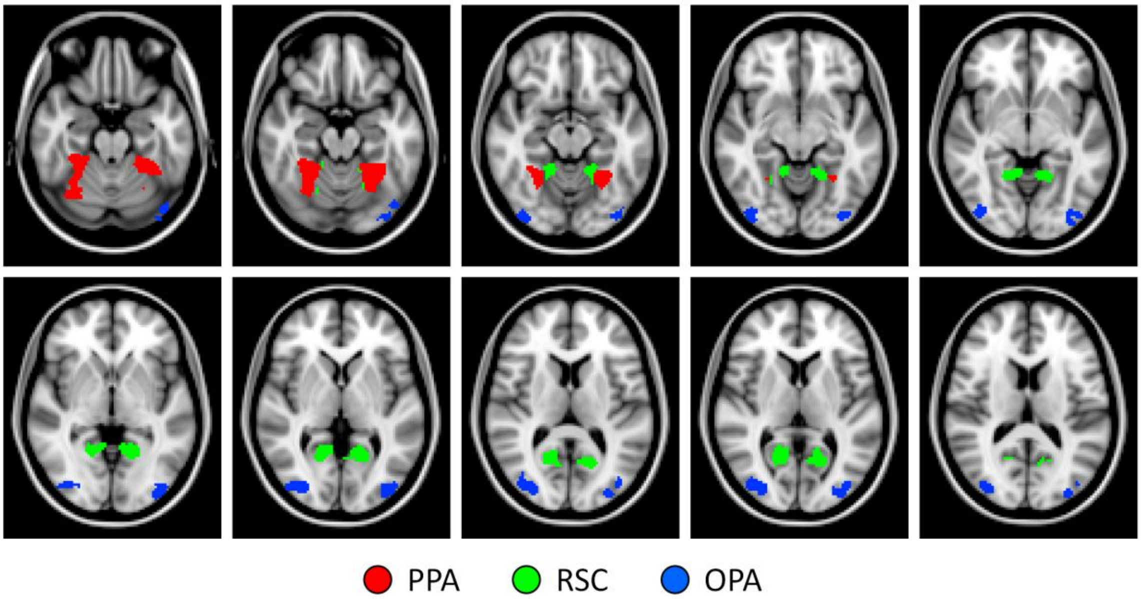


Figure 6.3. Illustration of the masks used for the fMRI analyses. Each mask comprises approximately 500 voxels (4000mm^3) in each hemisphere. Slices of MNI brain span the range from z = -22mm to z = +18mm in 4mm increments.

Table 6.1 Peak MNI mm co-ordinates, voxel counts, and thresholds of standard scene-selective clusters (PPA, RSC, OPA).

Region	Hemisphere	*x*	*y*	*z*	Voxel count	Threshold (Z)
PPA	L	-34	-46	-22	500	5.06
	R	26	-50	-18	500	5.59
RSC	L	-18	-52	2	500	4.63
	R	16	-58	6	502	4.79
OPA	L	-36	-90	2	500	5.14
	R	38	-82	4	499	5.03

Next, we measured patterns of response to different stimulus conditions in each experiment. Parameter estimates were generated for each condition in the experimental scans. The reliability of response patterns was tested using a leave-one-participant-out (LOPO) cross-validation paradigm (Shinkareva et al., 2008; Poldrack et al., 2009) in which parameter estimates were determined using a group analysis of all participants except one. This generated parameter estimates for each scene condition in each voxel. This LOPO process was repeated such that every participant was left out of a group analysis once. These data were then submitted to correlation-based pattern analyses (Haxby et al., 2001, 2014) implemented using the PyMVPA toolbox (http://www.pymvpa.org/; Hanke et al., 2009). Parameter estimates were normalised by subtracting the voxel-wise mean response across all experimental conditions per fold of the cross-validation (Haxby et al., 2001). For each iteration of the LOPO cross-validation, the normalized patterns of response to each stimulus condition were correlated between the group and the left-out participant. This allowed us to determine whether there are reliable patterns of response that are consistent across individual participants.

6.3.6 Semantic Model

We adapted the local semantic concept model proposed by (Greene & Oliva, 2009b) to determine the semantic similarity of the scenes. Objects within each of the scenes were manually segmented and labelled using the LabelMe toolbox (Russell et al., 2008). Objects were then re-labelled by one of 22 core object labels. These comprised all 16 labels employed by (Greene & Oliva, 2009b) (*sky, water, foliage, mountain, snow, rock, sand, animal, hill, fog, cloud, grass, dirt, manmade object, canyon,* and *road*), plus an additional 6 labels (*manmade structure, people, footpath / paved area, room interior, foodstuff,* and *vehicle*) necessary to fully describe the scenes within our stimulus set. Figure 6.4a illustrates this process for an example image. For each image a vector of 22 values was constructed where each value indicates the proportion of pixels within the image occupied by a given object label (Figure 6.4b). Each vector was then normalised to have a magnitude of 1. Finally, we constructed a correlation based similarity matrix from these vectors using a leave-one-image-out cross-validation procedure as described previously (Figure 6.4c).

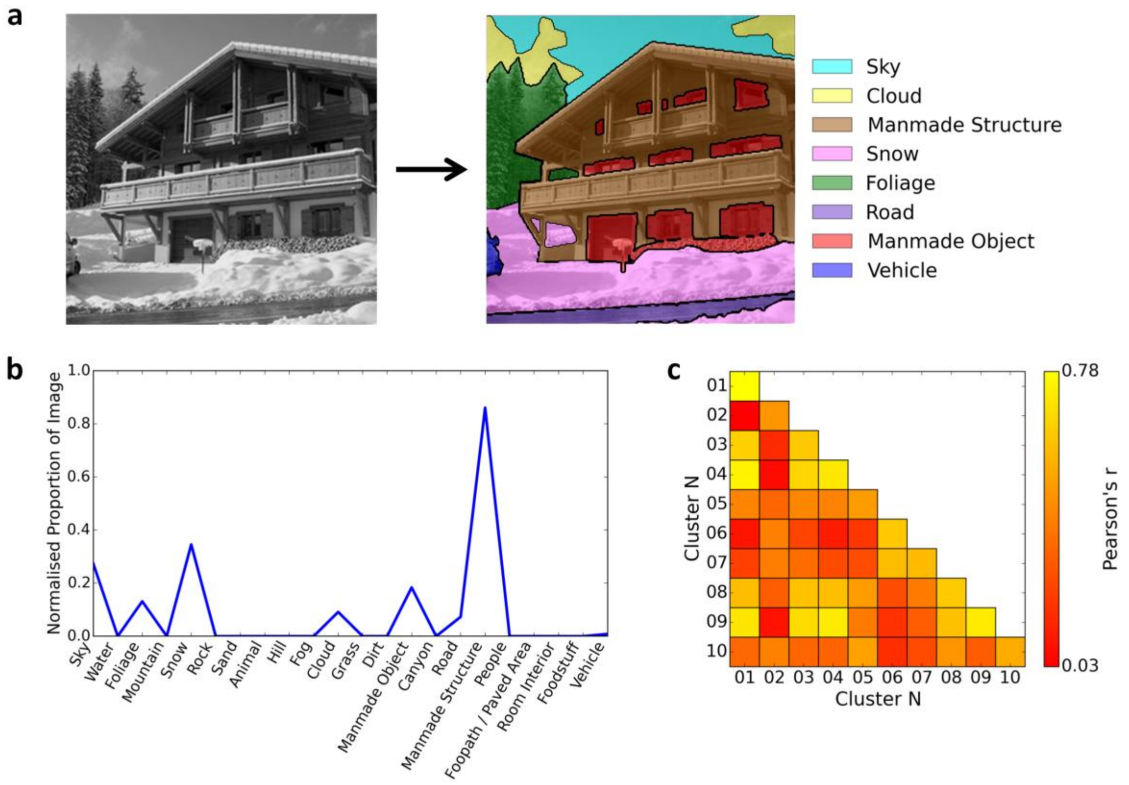


Figure 6.4. Local semantic concept model (Greene & Oliva, 2009b). (a) Objects within each of the images in the stimulus set were segmented and labelled using the LabelMe toolbox (Russell et al., 2008). Object labels were then reduced to a core set of 22 labels sufficient to describe the stimulus set. (b) For each image, a vector was calculated representing the proportion of pixels in the image occupied by each of the object labels. Vectors were normalised to have an overall magnitude of 1. (c) Group average similarity matrix calculated by correlating the vectors within and between clusters using a leave-one-image-out cross-validation scheme.

6.3.7 Behavioural Model

Participants completed a post-scan behavioural test, following a minimum delay of one week after the scan session in order to reduce bias by familiarity with the scenes. Written consent was obtained for all participants and the study was approved by the University of York Psychology Department Ethics Committee. Participants performed a card sorting task (Jenkins et al., 2011). Each participant was provided with a set of printed cards depicting one of four subsets of the scene set (60 images; 6 per cluster). Subsets were counterbalanced across participants. Participants were required to sort the cards into 10 stacks according to their perceptual similarity so that cards within a particular stack were ones that they perceived to all be similar to one another. The task was designed to allow participants as much freedom as possible to sort the cards however they wished. The precise definition of perceptual similarity was left deliberately vague so as to encourage participants to form their own interpretation. Card stacks were allowed to vary in size, and participants were allowed unlimited time to complete the task. In order to prevent the paradigm becoming a memory task, participants were required to stack cards next to one another so that they could always be seen.

Following the test, the number of cards from each of the scene clusters was counted for each of the card stacks. For each cluster a vector of 10 values was constructed representing the counts for that cluster across each of the card stacks. The lower-triangle of a similarity matrix was constructed by taking the dot-product of the vectors between each unique pairing of clusters, such that the element at position ***(i,j)*** represents the dot product between the vectors of the $\boldsymbol{i^{th}}$ and $\boldsymbol{j^{th}}$ scene clusters respectively. Values thus represent the frequency of co-occurrence between pairs of scene clusters across card stacks. This process is illustrated for an example subject in Figure 6.7a, and the group average similarity matrix is shown in Figure 6.7b.

6.3.8 Statistical Analyses

A Fisher's z-transform was applied to the correlation similarity matrices (GIST, MVPA, semantic) before further statistical analyses. A Bonferroni-Holm correction for multiple comparisons was applied across ROIs.

We first tested the MVPA and semantic models for their ability to decode the scene clusters. For each iteration of the cross-validation, we calculated the average within cluster (on-diagonal) and between cluster (off-diagonal) values of the correlations matrix. These values were then entered into a paired-samples t-test. If scene clusters can be discriminated, then significantly greater within than between-cluster correlations would be expected.

We also conducted a series of representational similarity analyses (RSAs; Kriegeskorte et al., 2008). Correlation matrices (GIST, MVPA, semantic) were averaged across iterations of the cross-validation, whilst the behavioural dot product matrices were averaged across subjects. Representational similarity was assessed by correlating the averaged similarity matrices between each of the models. The behavioural model only comprises the lower-triangle of the matrix, so only off-diagonal matrix elements were compared with this model. If any model is able to predict any other, a significant correlation would be expected between the respective similarity matrices. In order to investigate whether the local semantic concept model is able to explain any additional variance beyond the GIST model, we conducted a series of further RSAs using partial correlations. The MVPA correlation matrices were correlated with either the GIST or local semantic concept model whilst controlling for the effects of the other model. If the semantic model can explain any variance above and beyond the GIST model, a significant partial correlation between the semantic and MVPA models would be expected even when controlling for the GIST model. We also tested partial correlations between the behavioural and GIST / semantic models, and between the MVPA and GIST / behavioural models in the same manner.

6.4 Results

A data-driven analysis was used to define scene clusters based on their image properties (Figure 6.1). Examples of images in each cluster are shown in Figure 6.2. It is clear that these images do not reflect the typical scene categories commonly used in studies of scene perception. Next, we measured the pattern of neural response in each scene region (PPA, RSC, OPA) to the 10 different scene clusters using a blocked fMRI design. Figure 6.5 shows the normalised responses within each of the scene-selective regions (PPA, RSC, and OPA); red and blue colours indicate responses above and below the voxel-wise mean respectively.

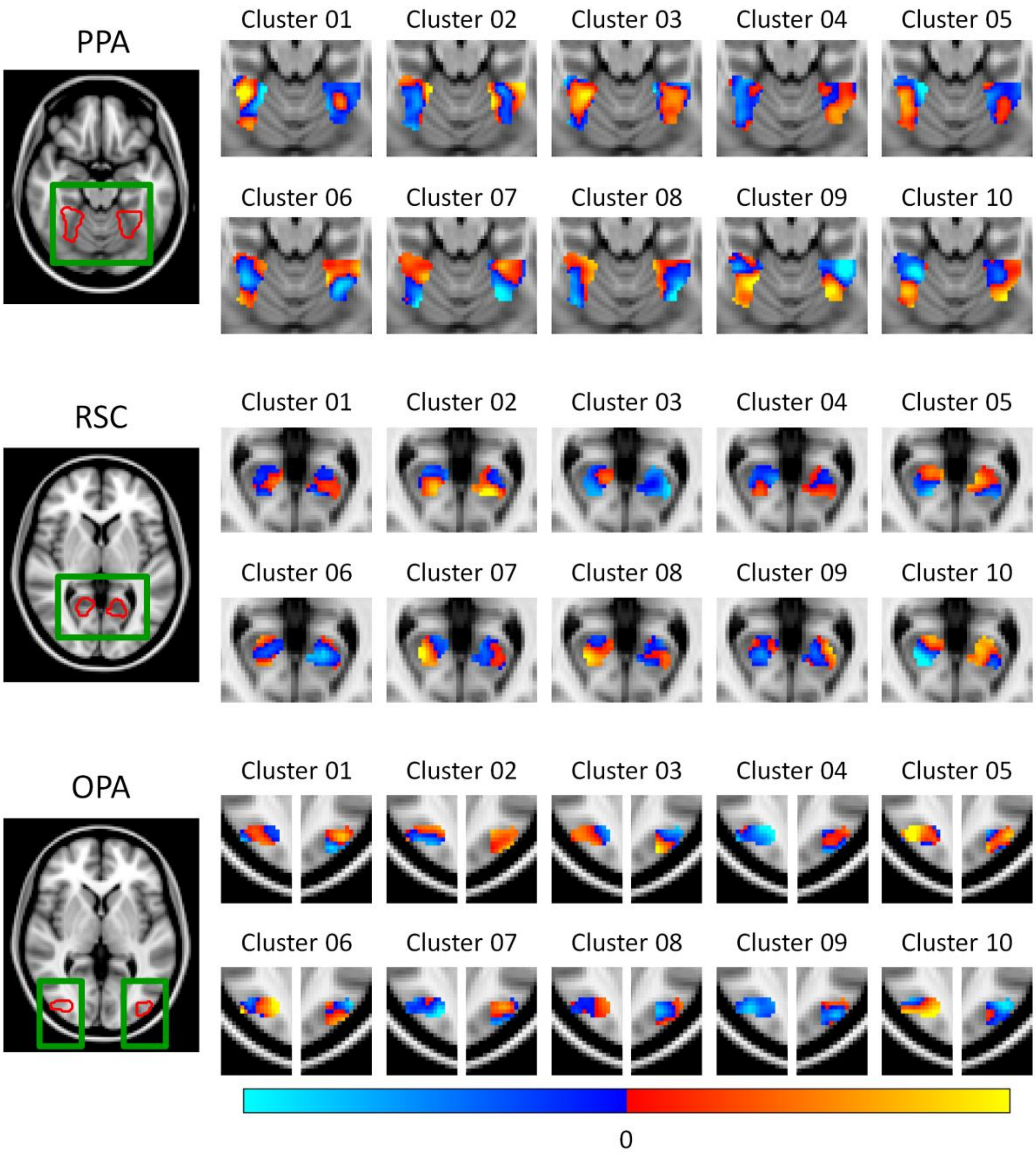


Figure 6.5. Group patterns of response restricted to each of the scene-selective regions (PPA, RSC, OPA). Responses are normalized by subtracting a voxel-wise mean across all conditions, such that red and blue colours indicate values above and below the mean response respectively.

Correlation-based MVPA (Haxby et al., 2001) was used to assess the reliability of these responses. Average correlation similarity matrices for each of the scene regions are shown in Figure 6.6a. We first assessed the ability of the MVPA to discriminate the scene clusters by comparing the within-cluster (on-diagonal) and between-cluster (off-diagonal) values of the correlation matrices. Figure 6.6b shows that there were significantly greater within-cluster than between-cluster correlations in the PPA (t(19) = 5.98, p < .001, Cohen's d = 1.34) and OPA (t(19) = 3.98, p = .002, Cohen's d = 0.89), but not in the RSC (t(19) = 0.10, p = .918, Cohen's d = 0.02). This shows that there are distinct neural responses in both PPA and OPA to the scene clusters defined by the data-driven method.

Next, we asked whether the similarity in the patterns of neural response to different scene clusters could be predicted by the similarity in the low-level image properties defined by the GIST descriptor. Using a representational similarity analysis (Kriegeskorte et al., 2008), we compared the correlation matrices for each region with the GIST correlation matrix. Results of these analyses are illustrated in Figure 6.6c and show that the image properties significantly correlated with neural responses in the PPA (r = .65, p < .001), but not the RSC (r = .21, p = .126) or OPA (r = .28, p = .081). It is possible that the correlation between image properties and neural responses in the PPA could be driven by the overall more positive on-diagonal than off-diagonal values in both the MVPA and GIST correlation matrices. To address this issue, we repeated our analysis using only the off-diagonal elements of each matrix. The correlation with the PPA remained statistically significant (r = .43, p = .008), while neither the correlations with the RSC (r = .26, p = .182) or OPA (r = -.04, p = .770) reached significance. This shows that the representational similarity structure of patterns of response in the PPA can be predicted by low-level image properties.

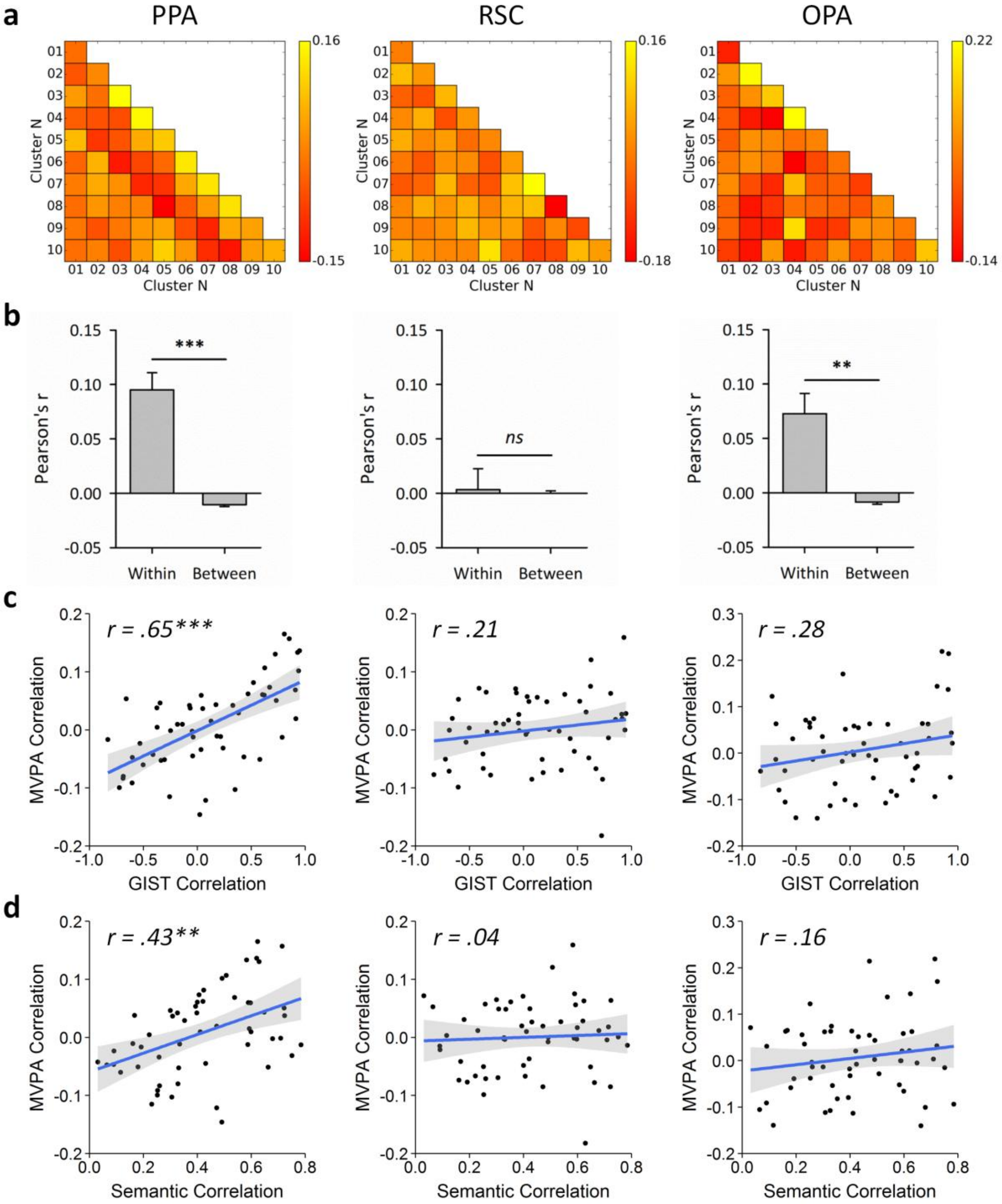


Figure 6.6. Main fMRI analyses for each scene region. (a) MVPA correlation matrices. (b) Discrimination of scene clusters by contrasting within over between cluster correlation values; error bars represent 1 SEM. Scatter plots show results of representational similarity analyses between the MVPA and: (c) the GIST and (d) local semantic concept correlation models; shaded regions represent 95% confidence intervals.

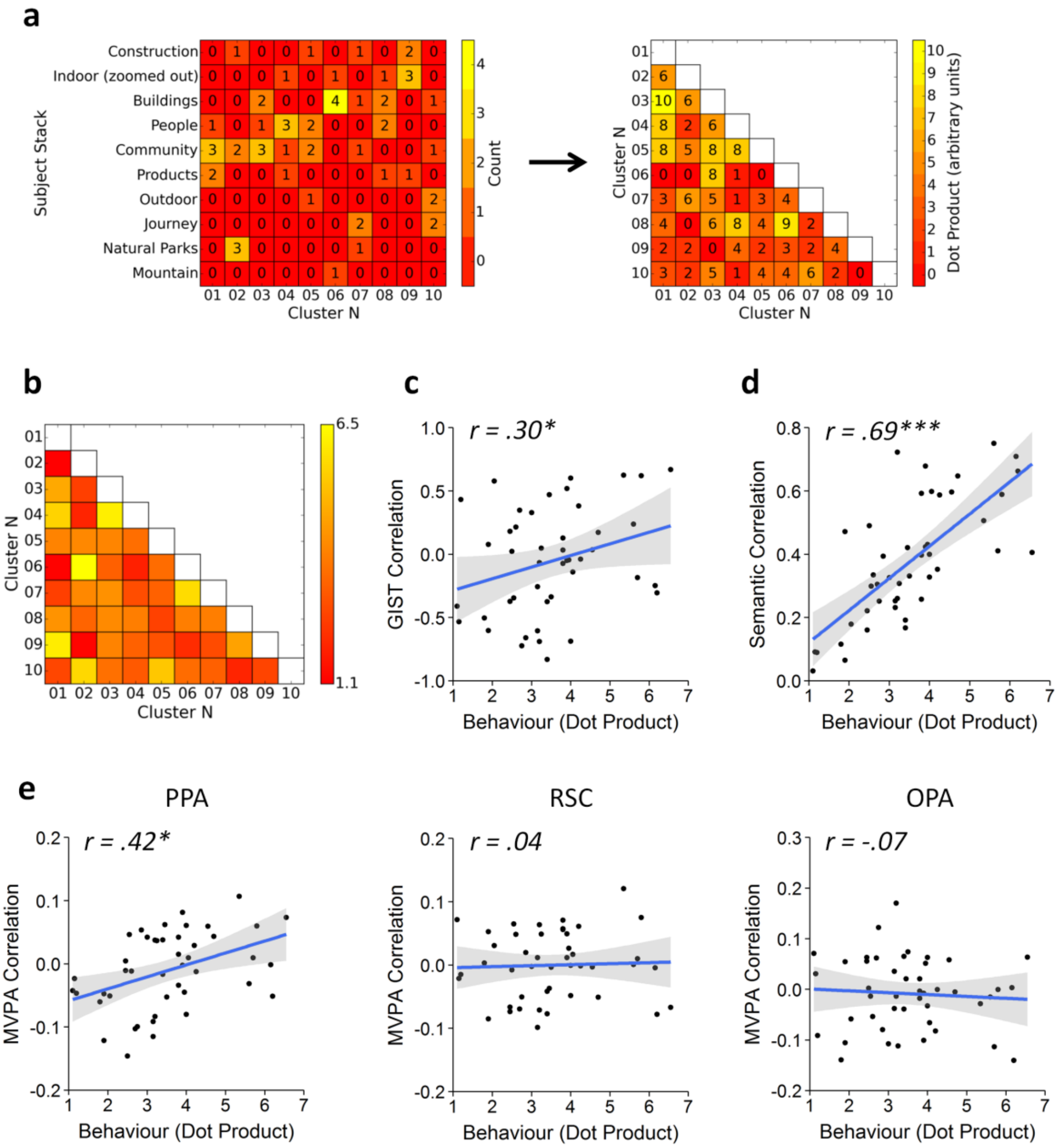
Although images in each cluster were selected on the basis of their visual properties, they also convey semantic information. For instance, scenes containing semantically similar objects also tend to be visually similar. To address this issue, the local semantic concept model (Greene & Oliva, 2009b) was used to test the semantic similarity of images within and between different clusters (Figure 6.4). To determine if each image cluster conveys distinct semantic information, we compared the within-cluster (on-diagonal) and between-cluster (off-diagonal) values of the correlation matrix. A paired-samples t-test revealed significantly higher within- than between-cluster correlations (t(23) = 12.67, *p* < .001, Cohen's d = 2.56), indicating that clusters could be discriminated based on semantic properties. We next determined the representational similarity between the local semantic properties and the image properties given by the GIST analysis. We found a significant positive correlation between semantic and image properties (*r* = .48, *p* < .001). Repeating this analysis with only the off-diagonal elements revealed a reduced correlation, but one that was nevertheless borderline significant (*r* = .29, *p* = .050).

Next, we asked whether the local semantic properties could predict the patterns of fMRI response in scene-selective regions by correlating the respective correlation matrices. Semantic properties significantly correlated with neural responses in the PPA (*r* = .43, *p* = .003), but not the RSC (*r* = .04, *p* = .788) or OPA (*r* = .16, *p* = .507). These results are illustrated in Figure 6.6d. When we repeated our analyses using only the off-diagonal elements of the matrices no significant correlations were found for any region (PPA: *r* = .28, *p* = .193; RSC: *r* = .11, *p* = .942; OPA: *r* = .04, *p* = .942). However, as the semantic and GIST models are themselves correlated it remains unclear whether the semantic model is able to explain significantly more variance in the PPA data above and beyond that already explained by the GIST model. To test this, we repeated our analyses for the PPA region using partial correlation to control for the effect of one or the other model. A significant partial correlation was observed between neural responses and the GIST model while controlling for the semantic model (*r* = .56, *p* < .001). However, the partial correlation with the semantic model while controlling for the GIST did not reach significance (*r* = .18, *p* = .196). A similar pattern of results was observed when restricting the analysis to only the off-diagonal elements; both when correlating neural responses with the GIST while controlling for the semantic model (*r* = .39, *p* = .009), and when correlating neural responses with the semantic whilst controlling for the GIST model (*r* = .17, *p* = .258). Thus the neural responses were primarily predicted by the GIST model, whilst the semantic model did not significantly predict any additional variance.

Finally, we tested how each of our models compared to human behaviour. Participants completed a card-sorting task in which scenes were sorted into distinct stacks according to their perceptual similarity (Jenkins et al., 2011). A similarity matrix was constructed by examining the co-occurrence of each possible pairing of scene clusters across each of the subject's card stacks. This was calculated by defining a vector for each scene cluster denoting the counts across each of the card stacks, and taking the dot product between each pairwise combination of vectors (Figure 6.7a). The average dot product similarity matrix is shown in Figure 6.7b. We first tested the representational similarity with the GIST and semantic models. Because the behavioural similarity matrix contains only the lower-triangle, only off-diagonal elements were compared between models; results of these analyses are shown in Figure 6.7c-d. A significant correlation was found between the behavioural responses and both the GIST (*r* = .30, *p* = .045) and semantic models (*r* = .69, *p* < .001). Repeating these analyses as partial correlations revealed a significant partial correlation between the behavioural and semantic models while controlling for the GIST model (r = 0.66, *p* < .001). However, the partial correlation between the behavioural and GIST models while controlling for the semantic model failed to reach significance (*r* = 0.14, *p* = .361). Thus the behavioural responses were primarily predicted by the semantic model.

We next asked whether the behavioural responses could predict patterns of neural response (Figure 6.7e). Behavioural responses significantly correlated with neural responses in the PPA (*r* = .42, *p* = .012), but not the RSC (*r* = .04, *p* > .999) or OPA (*r* = -.07, *p* > .999). In order to compare the unique contributions of the behavioural and visual models to the PPA response, we repeated our analyses using partial correlations. Significant correlations were observed both when comparing the neural response with the GIST whilst controlling for the behavioural model (*r* = .36, *p* = .017), and comparing the neural response with the behavioural model whilst controlling for the GIST (*r* = .34, *p*

= .024). Thus, the visual and behavioural models account for relatively distinct components of the variance in the PPA response.



	01	02	03	04	05	06	07	08	09	10
Construction	0	1	0	0	1	0	1	0	2	0
Indoor (zoomed out)	0	0	0	1	0	1	0	1	3	0
Buildings	0	0	2	0	0	4	1	2	0	1
People	1	0	1	3	2	0	0	2	0	0
Community	3	2	3	1	2	0	1	0	0	1
Products	2	0	0	1	0	0	0	1	1	0
Outdoor	0	0	0	0	1	0	0	0	0	2
Journey	0	0	0	0	0	0	2	0	0	2
Natural Parks	0	3	0	0	0	0	1	0	0	0
Mountain	0	0	0	0	0	1	0	0	0	0

	01	02	03	04	05	06	07	08	09	10
01										
02	6									
03	10	6								
04	8	2	6							
05	8	5	8	8						
06	0	0	8	1	0					
07	3	6	5	1	3	4				
08	4	0	6	8	4	9	2			
09	2	2	0	4	2	3	2	4		
10	3	2	5	1	4	4	6	2	0	

Figure 6.7. Behavioural experiment method and results. (a) Illustration of the analysis procedure for an example subject. A matrix of counts (left) was generated for each of the scene clusters (columns) against each of the subject's card stacks (rows). The card stack labels were generated by the subject themselves. The lower triangle of a similarity matrix (right) was then constructed by calculating the dot-product between each pairwise combination of columns in the counts matrix. The group average dot product similarity matrix (b) was then compared against the GIST (c), local semantic concept (d), and MVPA models (e) in a series of representational similarity analyses. Shaded regions on scatterplots indicate 95% confidence intervals.

6.5 Discussion

The aim of this study was to explore the functional organization of scene-selective regions in the human brain using a wholly data-driven method. Clusters of scenes were defined objectively by their image properties. Our results show that in the scene-selective PPA there are 1) distinct patterns of response to each scene cluster, 2) that the similarity of neural responses to different scene clusters is well explained by the similarity of the corresponding visual descriptors, and 3) that the semantic properties of the different scene clusters do not explain any additional variance in the neural responses beyond that explained by the visual descriptors. Together, these results demonstrate a clear link between patterns of response in scene-selective regions and low-level image properties.

Patterns of response in PPA are closely related to visual properties

We demonstrated that visually defined clusters generated distinct patterns of response in the PPA. Furthermore, responses in the PPA showed a similar representational structure to that predicted by a low-level image-based model: the more similar the visual properties of each cluster, the more similar the pattern of neural response. These results fundamentally extend those of previous experiments by demonstrating sensitivity to the visual properties of the scene, independent of prescribed scene categories.

Previous studies have revealed distinct patterns of neural response to different scene categories within scene-selective cortices (Walther et al., 2009, 2011). Such findings have been taken to suggest a functional organisation that tracks the high-level categorical aspects of scenes, and is at least partially independent of image properties. In contrast, later studies have suggested that effects of category may be better explained by visual properties of scenes, such as openness (Kravitz et al., 2011; Park et al., 2011) or distance (Amit et al., 2012; Park et al., 2015). However, an important limitation in previous studies is the fact that the choice of stimulus conditions was determined by the experimenters. In each case, although the manipulated factors clearly influence the neural response, they need not correspond to fundamental organizing principles.

The scene clusters used in the current study are essentially arbitrary in terms of the scene categories used in classic designs. Our results demonstrate that scene category need not be considered the dominant organising principle of scene-selective regions. More parsimoniously, the effects of manipulations of scene category on patterns of neural response, seen in many earlier experiments, are likely driven by systematic variation in the underlying scenes' image statistics. Indeed, even in the current study, when grouping images according to objective visual descriptors, a statistically significant relationship between semantic content and visual properties remains, indicating that these characteristics cannot be considered entirely independent in natural images. For instance, scene cluster 6 is marked by images with a strong horizontal component across the middle of the image, frequently manifested as outdoor scenes with a strong horizon line. Although this is a visual distinction, it also means scenes are frequently associated with labels such as "sky" and "cloud", but less so labels such as "vehicle" or "animal". Critically, however, the semantic properties of each cluster did not account for additional variance in representational similarity of the neural response after controlling for visual properties. In contrast, the visual properties of each cluster predicted representational similarity in patterns of neural response after controlling for semantic properties.

Behavioural classification is better explained by local semantic object information

How do visually-organised patterns of responses contribute to perception and categorisation? Many previous studies have demonstrated our ability to categorise scenes (Schyns & Oliva, 1994; Oliva & Schyns, 1997; Greene & Oliva, 2006; Xiao et al., 2010; Ehinger et al., 2011). However, like the earlier neuroimaging research, these studies typically rely on tasks that are constrained by the choice of categories. Here, we used a card-sorting task that allowed participants a high degree of freedom in choosing how to group the scenes used in the fMRI experiment (Jenkins et al., 2011). If there were a direct link between neural responses and perceptual decisions, we might expect a linear relationship between representational similarity associated with fMRI responses to scene clusters and participants' behavioural classification of the same items. Indeed, both visual and behavioural models were found to significantly predict the representational similarity of neural responses in the PPA, and each explained relatively independent components of the variance. Yet, while the visual descriptor model provided the better account of patterns of neural response, the local semantic model explained the most variance in participants' unconstrained behavioural classification of the stimuli, suggesting a partial dissociation between the mechanisms driving patterns of neural response in PPA and those responsible for categorical perception.

Patterns of response in other scene selective regions

Our most significant positive findings concern the PPA, but we found interesting differences in the response profiles between other scene-selective regions (OPA, RSC). RSC responses failed to discriminate the visually-defined scene clusters, and the representational similarity structure in this region was not predicted by any of the models tested. Previous studies have identified complimentary but distinct roles for the PPA and RSC (Byrne et al., 2007; Epstein & Higgins, 2007; Epstein et al., 2007a; Epstein, 2008; Park & Chun, 2009; Marchette et al., 2014, 2015), with the PPA proposed to be involved in processing the spatial features in the immediate visual environment, while the RSC focuses more on integrating the scene within the wider spatial environment, and in mediating translations between egocentric and allocentric representations. Our results are consistent with this view since the GIST descriptor captures the critical (image-based, egocentric) spatial variables that are thought to underlie scene perception (Torralba & Oliva, 2003), but may be less directly relevant to the more abstract representations required for the integration of scenes within the wider environment and the extraction of allocentric information.

A somewhat different pattern of results was observed in the OPA. Although showing distinct and reliable patterns of response to different scene clusters, these only weakly maintained the representational similarity predicted by the GIST descriptor. Furthermore, neither local semantic concept nor behavioural models predicted the representational similarity structure seen in OPA responses. These findings are consistent with a proposal for a hierarchical network of scene processing in which more posterior regions such as the OPA are sensitive to visual properties in scenes, but are perhaps less selective for spatially diagnostic features than more anterior regions such as the PPA (Kravitz et al., 2011; Park et al., 2011; Dilks et al., 2013).

Conclusion

In conclusion, we describe a method for data-driven clustering of scenes based on their image properties. This overcomes limitations of more traditional experimental designs in which scene stimuli are subjectively allocated to predefined categories. We demonstrate that scene selective regions, in particular the PPA, display a clear sensitivity to the low-level visual properties of scenes, independent of prescribed scene categories. Local semantic properties of the scene are correlated with visual properties, but fail to explain additional variance. However, behavioural classification of the scenes was better explained in terms of local semantic properties than image properties. Overall the results underscore the importance of visual features in functional responses of scene-selective regions of the human brain, and suggest that scene category need not be the dominant organising principle.

Chapter 7 – General Discussion

Human observers are able to rapidly perceive and extract the key spatial components of visual scenes – the so called scene *gist* (Oliva & Torralba, 2001; Torralba & Oliva, 2003). Indeed, observers are able to reliably perceive and categorise scenes even when images are presented rapidly (Potter, 1975; Greene & Oliva, 2009a) or under degraded visual conditions (Torralba, 2009). Recent neuroimaging studies have identified a number of cortical regions responsive to images of visual scenes, which are thought to underlie our ability to perceive scenes. These regions include the parahippocampal place area (PPA; Aguirre & D'Esposito, 1997; Epstein & Kanwisher, 1998), retrosplenial complex (RSC; Maguire, 2001; Vann et al., 2009), and the transverse occipital sulcus / occipital place area (TOS / OPA; Dilks et al., 2013).

Whilst the existence of these regions is well established, the precise stimulus dimensions underlying their functional organisation remain controversial. Some accounts have argued for relatively high-level accounts based upon categorical or semantic properties of the scene, based on the finding that distinct patterns of neural response can be observed to different semantic categories of scene (Walther et al., 2009, 2011). However, other studies have argued for an organisation based on more mid-level spatial envelope properties of scenes (Kravitz et al., 2011; Park et al., 2011). Other studies still have argued for even lower-level accounts based on biases for visual features, such as those for spatial frequency (Rajimehr et al., 2011), orientation (Nasr & Tootell, 2012), rectilinearity (Nasr et al., 2014), and retinotopy (Malach et al., 2002; Arcaro et al., 2009). A complication in this debate has been that many of these features are themselves correlated (Lescroart et al., 2015), thus making it difficult to separate out the effects of any one account over the others. For instance, visual features have been shown to be predictive of both the spatial content and semantic category of scenes (Oliva & Torralba, 2001; Torralba & Oliva, 2003).

Thus, it remains unclear from previous research what the relative contributions of high- and low-level properties of scenes are to the function of scene selective regions.

Furthermore, although previous studies have identified low-level visual biases within these regions, many of them have employed univariate analyses and so it remains unclear to what extent such properties may be reflected in distributed *patterns* of neural response. Therefore, the primary aims of this thesis were:

- To determine if low-level visual properties of scenes can predict patterns of neural response within scene-selective visual cortices.
- To test the relative contributions of low-level visual information against high-level semantic or categorical information to the pattern of response.
- To determine the contributions of specific low-level visual properties (i.e. spatial frequency, orientation, and retinotopy) to the neural response.
- To explore whether alternatives to more traditional categorical accounts might more parsimoniously explain the function of scene selective cortices.

To this end, a series of neuroimaging experiments employing fMRI in conjunction with multi-variate pattern analysis (MVPA) were conducted to investigate these possibilities.

7.1 The representation of scenes in the brain

If neural representations of scenes are related to the visual properties of the image, then one would expect that neural responses could be predicted by a model of the visual features of scenes. Chapter 3 presented two fMRI experiments that measured the patterns of neural response to different categories of scene; city, indoor, and natural scenes in the first experiment, and coast, forest, and mountain scenes in the second experiment. The low-level visual properties of the scenes were measured by the GIST descriptor (Oliva & Torralba, 2001). The GIST descriptor is designed to capture the critical spatial features of scenes, and the key visual features measured by the GIST (spatial frequency, orientation, and retinotopy) map on to known tuning properties of neurons in visual cortex (Wandell & Winawer, 2011), thus providing a neurologically plausible model of visual statistics. Representational similarity analyses showed that, in both experiments, the GIST descriptor was able to predict the relative similarity between neural response patterns to the different scene categories.

It should be noted that the GIST descriptor is not the only image descriptor available – other popular algorithms include the HMAX model (Riesenhuber & Poggio, 1999), HOG descriptor (Dalal & Triggs, 2005), and SIFT descriptor (Lowe, 2004). However, it was beyond the scope of this thesis to provide a comprehensive comparison of the different visual descriptors, and thus the GIST was the only model tested. Nevertheless, the GIST remains a good choice of model for a number of reasons. Firstly, the GIST is theoretically motivated for scene processing, which many other models are not, and indeed the GIST has been demonstrated to successfully discriminate scenes computationally (Oliva & Torralba, 2001). Secondly, the GIST model is neurologically plausible and thus overcomes the limitations of previous experiments that used less plausible image models, such as pixel correlations (e.g. Walther et al., 2009). Finally, the simplicity of the GIST model and the relatively coarse-scale at which it samples the image may provide a good correspondence to the relatively coarse-scale at which fMRI samples the neural response. This contrasts with other models such as the HMAX which, although neurologically plausible, nevertheless aim to provide a model of the image closer to the resolution of individual neurons.

Although these results do lend support to the importance of low-level visual features, they do nevertheless remain correlational. The use of distinct scene categories in the design of these experiments necessarily confounded the effects of visual properties with those of scene category. To address this, Chapter 4 presented a further two fMRI experiments in which visual properties of the stimuli were directly manipulated. In the first of these, scenes were filtered between horizontal and vertical orientation content, whilst the second of these filtered the scenes between low and high spatial frequency content. Importantly, both experiments employed a 2x2 design in which the relevant levels of filtering were applied across two scene categories (indoor and natural), thus allowing a direct comparison of the effects of visual filter and category. In the first of these experiments, little to no effect of the orientation content was seen upon the neural response patterns. This does not necessarily mean that orientation is not represented by scene selective regions (indeed other recent reports have suggested orientation biases within these regions, e.g. Nasr & Tootell, 2012; Nasr et al., 2014), but may suggest that these properties are not represented in the distributed patterns of response measured by the MVPA technique. It should also be noted that Nasr & Tootell (2012) reported a bias of cardinal over oblique orientations, which would not have coincided with the horizontal and vertical orientation distinction used in this experiment. By contrast, the second of these experiments revealed a highly significant effect of spatial frequency in all regions of interest. Furthermore, the magnitude of this effect was found to be significantly greater than the effect of category in the both the overall scene selective region and the OPA, whilst the PPA and RSC showed more comparable effects of spatial frequency and category. By directly manipulating the visual content of the scenes, this approach provides stronger evidence for the influence of visual properties upon the functional organisation of scene selective regions. Although some residual effects of scene category were observed in these experiments, this does not necessarily have to reflect direct effects of semantic category. For instance, although the scene categories did differ in terms of semantics, they also differed in many other visual features beyond the ones manipulated by the filters. Thus, effects of category here can be thought of as reflecting all remaining stimulus dimensions after the effect of the filter has been accounted for, and therefore although they could be attributable to high-level semantic differences, they could equally well reflect additional sensitivity to the many other visual features that differed between the categories.

Chapter 5 presented an fMRI experiment testing the contribution of visual properties of scenes to the corresponding neural responses under conditions where perception of the scene content was disrupted. Fourier phase scrambling was applied to images of scenes from 5 categories (city, coast, forest, indoor, and mountain) either globally (across the whole image) or locally (within windows of a 4x4 grid across the image). Both methods of scrambling significantly impaired subjects' ability to recognise the scenes. Despite the fact that the scrambling severely impaired scene perception, it was found that scene categories could nevertheless be discriminated from the neural response patterns to the locally scrambled scenes in the PPA and OPA. Furthermore, a series of representational similarity analyses demonstrated that PPA and OPA responses to intact scenes were similar to those to the scrambled scenes. Again, the direct manipulation of the visual content of the images here provides strong evidence for the role of visual properties in the functional responses of scene regions. The fact that effects of scene category were observed even when the perception of these categories was clearly impaired would suggest that representations are more closely tied to visual features of the images that differ reliably between the categories, rather than the semantic category itself. In the case of the PPA, it was also found that responses to the intact scenes were more similar to the locally scrambled than the globally scrambled scenes, suggesting some sensitivity to the retinotopic distribution of the visual features.

The experiments discussed so far all contain one common design choice – that scenes are selected from pre-defined categories. This approach to stimulus selection is also extremely common within the literature, almost universal. However, such design decisions can also be problematic as the selection of these categories is necessarily subjective and biased by experimenter choices. Selection of these categories may not be justified by the data, i.e. there is not necessarily any reason to assume that "scene category" should be the most natural feature along which neural responses should be expected to vary. An experiment that employs conditions following scene categories may well demonstrate a significant effect of those categories, but it could not very well have demonstrated an effect of anything else if the design does not permit such alternative hypotheses to be investigated, and therefore risks obscuring other potentially simpler accounts of neural function. To address these concerns, Chapter 6 presented one final fMRI experiment in which scenes were selected objectively based upon their visual properties, as measured by the GIST descriptor, thereby avoiding experimenter bias in choosing the stimulus conditions. The resulting clusters of scenes were essentially arbitrary in terms of semantic category, but did differ reliably in their low-level visual properties. It was found that patterns of neural response not only successfully discriminated the scene clusters, but also modelled a similar representational similarity to that predicted by the GIST. Furthermore, neural responses were better predicted by the GIST descriptor than by a model of the semantic object properties of the scene. These results therefore provide strong support for the importance of visual properties in determining responses of scene selective regions, whilst arguing against scene category being the dominant organising principle of such regions. Interestingly, the semantic object model did prove a better model of human behaviour than the GIST descriptor. This suggests a possible dissociation between human behaviour and neural responses, with behaviour more closely tracking semantic / object properties, whilst neural responses more closely track visual properties. However, a significant relationship was seen between behavioural and neural responses, indicating this dissociation is only partial.

Thus, the evidence presented in this thesis would argue against an organisation of scene selective regions based upon categorical principles. Instead, the results would seem to favour an account in which neural response patterns are closely tied to the low-level visual properties of the stimuli. This conclusion would be consistent with previous studies employing univariate analyses that have reported low-level visual biases (Malach et al., 2002; Arcaro et al., 2009; Rajimehr et al., 2011; Nasr & Tootell, 2012; Kauffmann et al., 2014; Nasr et al., 2014; Silson et al., 2015). Whilst effects of scene category may be evident (Walther et al., 2009, 2011), it seems likely that these are largely accounted for by visual features that are known to differ reliably between scenes (Oliva & Torralba, 2001) rather than the semantic category *per se*. One key problem for more traditional accounts is explaining how image-based representations in early visual cortices are transformed into semantic or categorical representations in high-level visual cortices (Andrews et al., 2015). The results presented in this thesis suggest that such an explanation may in fact not be necessary; apparently high-level responses to specific stimulus classes can be explained by sensitivity to low-level visual features that are predictive of those classes.

In light of these results, one open question is: what role is there for non-visual properties in the function of scene regions? It seems possible, if not probable, that functional responses would be influenced by top-down neural feedback, and indeed studies have noted modulation of responses in ventral-temporal visual cortex by attentional and task demands (Harel et al., 2014; Kay et al., 2015). It is also possible that responses could be influenced by lateral connections, for instance in the form of cross-modal input (Wolbers et al., 2011). Nevertheless the evidence presented in this thesis argues strongly in favour of a dominant role of visual properties, but importantly this does not necessarily discount some additional influences of other non-visual properties.

It should be noted that the evidence presented here does not suggest that scene-selective regions are not truly responsive to scenes. Instead, it suggests a possible mechanism by which such scene-selectivity may arise. Rather than understanding scene-selective regions as responding to scenes as a category or a semantic concept, it may be that they respond to a combination of low-level visual features (and possibly some non-visual features) that are themselves predictive of scene content. Indeed, visual features such as those discussed here have been noted to reliably capture the key spatial geometries of scenes (Oliva & Torralba, 2001; Torralba & Oliva, 2003). This conclusion would be consistent with Op de Beeck et al.'s (2008) model, in which localised selectivity for specific stimulus classes is proposed to arise from the interaction of multiple lower-level biases mapped across the cortex and which are themselves predictive of each stimulus class.

7.2 Functional dissociations between scene regions

In all experiments, the three core regions of the scene network (PPA, RSC, and OPA) were identified. This allowed examination of the similarities and differences in response profiles between each of the regions. In all experiments, the PPA consistently displayed a clear sensitivity to the low-level visual properties of the scenes. The PPA has been proposed to be implicated in the extraction and processing of the local spatial geometries of scenes (Epstein, 2008). Given that low-level visual properties closely relate to the spatial structure of scenes (Oliva & Torralba, 2001), the sensitivity of the PPA to such features could therefore support its role in such processes.

In contrast, responses of the RSC appeared more variable. Some commonalities between responses and visual properties were observed. For instance, responses could be predicted by the GIST descriptor (Chapter 3), and were significantly modulated by the spatial frequency content of the image (Chapter 4). However, RSC responses failed to discriminate the different categories of scenes in scrambled images, and little similarity was seen between responses to intact and scrambled scenes (Chapter 5). Furthermore, RSC responses also failed to discriminate the objectively selected scene clusters (Chapter 6). This would suggest some degree of sensitivity of the RSC to the visual features of scenes, but one that is not as ubiquitous as that which is observed for the PPA, for instance. The RSC has been proposed to play a complementary role to the PPA, but one which is nevertheless distinct and focuses on more navigationally relevant processes such as locating the immediate visual scene within the wider spatial environment (Epstein, 2008; Vann et al., 2009). It is possible that the low-level visual content of the image is less important to these higher-level, navigationally relevant processes and thus a partial degree of independence from visual features might be expected.

Much like the PPA, responses in the OPA consistently displayed sensitivity to the visual features of scenes. In some cases, the magnitude of this effect was quite large – for instance the OPA was much more strongly modulated by the spatial frequency content than the category information of scenes (Chapter 4). However, some functional differences from the PPA were also observed. For instance, whilst both the PPA and OPA successfully discriminated responses to the objectively defined scene clusters, only the PPA responses were also predicted by the representational similarity of the GIST descriptor (Chapter 6). In comparison to the PPA and RSC, much less is known about the functional role of the OPA within the scene processing network. Dilks et al. (2013) propose that the OPA may represent an early stage within a hierarchical scene processing network, analogous to proposed roles for the occipital face area in the face processing network (Haxby et al., 2002). The results presented here suggest that whilst the OPA may maintain sensitivity to the visual features of scenes, it may be less concerned with representing those features in terms of the critical spatial dimensions of scenes than the PPA is. Such a conclusion would therefore be consistent with Dilks et al.'s proposal. Alternatively, Silson et al. (2015) note upper and lower visual field biases in the PPA and OPA respectively, and suggest a complimentary role for the two regions, with each representing a continuation of the lateral / ventral divide of regions in early visual cortex (e.g. V2v/V2d, V3v/V3d). Such biases could potentially contribute to the results observed here. For instance, differences in the (non-)stationarity of the visual statistics between scenes are often observed primarily in the upper sections of the image, e.g. by the presence or absence of sky between different types of scene (Torralba & Oliva, 2003). Upper and lower visual field biases might therefore be expected to produce differential responses in PPA and OPA regions. Nevertheless, it seems unlikely that visual field biases would be able to wholly explain the observed differences in response between the two regions, and so the existence of some additional functional distinctions between the regions seems likely.

One potential avenue of future research that may help resolve the functional relationships between scene regions would be to further examine their connectivity with one another. In particular, methods with high temporal resolution such as MEG may be able to resolve the temporal dynamics of connectivity within the scene processing network. It is conceivable that different stimulus features could exert different influences on the functional response across the course of the timeseries. For instance, a number of studies using MEG have reported that responses in occipital and ventral-temporal visual cortices represent increasingly complex stimulus features of visual objects over the progression of the timecourse in the first few hundreds of milliseconds after stimulus onset (Carlson et al., 2013; Cichy et al., 2014; Clarke et al., 2015). It is therefore possible that visual features of scenes may drive functional responses and connectivity within the scene network most strongly early in the timecourse, whilst a greater role for other non-visual features could be seen later in the timecourse.

7.3 Properties of the neural response patterns

It is frequently stated that the neural patterns informative to MVPA exist at a fine spatial scale and are largely idiosyncratic to each individual subject (Haxby et al., 2014). However, other studies have argued that patterns may instead be organised at a coarser spatial scale (Op de Beeck, 2010; Freeman et al., 2011). Furthermore, a number of studies have successfully employed cross-participant pattern analyses (Shinkareva et al., 2008, 2011; Poldrack et al., 2009), which would argue against purely idiosyncratic response patterns.

The fMRI experiments presented in Chapter 3 of this thesis reported the results of pattern analyses for both spatially smoothed and unsmoothed data, and for both individual-participant (cross-validating across odd and even stimulus blocks) and cross-participant analyses (using a leave-one-participant-out cross-validation scheme). It was found that spatially smoothing the data had little to no detrimental effect upon either the decoding or representational similarity analyses, and if anything spatial smoothing lead to a small benefit. Similarly, results of the cross-participant analyses appeared comparable to or slightly better than those of the individual-participant analyses, especially when performed upon spatially smoothed data. In all further fMRI experiments, analyses were conducted on spatially smoothed data using cross-participant analyses, with no clear detrimental effects. Thus, counter to common assumptions, the results presented here would argue for a functional organisation of scene selective regions based upon coarse-scale patterns of response that display at least some degree of commonality across subjects. It should be noted that this does not preclude the possibility that there may exist further pattern information that is fine-scale and / or idiosyncratic, but it does appear that if such information is present then it was not necessary to observe the significant differences between the patterns of response, or to measure the representational similarity of the patterns, as has been reported in this thesis. Nevertheless, it is possible that fine-scale and / or idiosyncratic patterns may be more prevalent in other brain regions, or for other experimental designs. For instance, an event-related design in which responses are measured independently for individual images might be expected to yield pattern information that is somewhat different to that given by measuring relatively generic responses across blocks of many images. Further research investigating the spatial scale and idiosyncrasy of neural response patterns in different brain regions and using different experimental methods may therefore be required to fully resolve this issue.

7.4 Conclusion

This thesis aimed to investigate the role of low-level visual properties in the representation of scenes in the brain. A series of fMRI experiments in conjunction with multi-variate pattern analyses revealed a clear sensitivity of scene-selective regions to the visual content of images, both in terms of responses being predicted by a model of low-level visual features (GIST), and in terms of responses being modulated by direct manipulation of such properties. This would suggest that scene selectivity in the brain may, at least in part, arise from multiple co-occurring biases for low-level visual features that are themselves predictive of scene content. Although all of the scene regions tested (PPA, RSC, and OPA) showed some degree of sensitivity to the visual content of scenes, there were nevertheless some differences between each of their response profiles. The PPA showed a clear sensitivity to the visual features of scenes, possibly supporting its proposed role in extracting local spatial geometries of visual scenes. The RSC also displayed some sensitivity to visual features, but less so than the PPA, which is possibly consistent with its proposed role in higher-level, more navigationally relevant aspects of scene processing. The OPA displayed clear sensitivity to the visual content of scenes, but appeared less concerned with representing such information in terms of the critical spatial geometries of scenes than the PPA was. This may suggest a role for the OPA as an early region within a hierarchical scene processing network, but this remains uncertain as literature on the response properties of the OPA is currently lacking. Finally, by using spatially smoothed data in conjunction with cross-participant analyses, these experiments demonstrated that pattern information represented in scene selective cortices can be measured in a coarse-scale manner that shows commonality across subjects. Taken together, these results therefore provide a significant contribution to the literature by demonstrating selectivity for low-level visual features of images in high-level, scene-selective visual cortices of the human brain.

Appendices

A.1 Supplementary Figures

A.1.1 Chapter 3

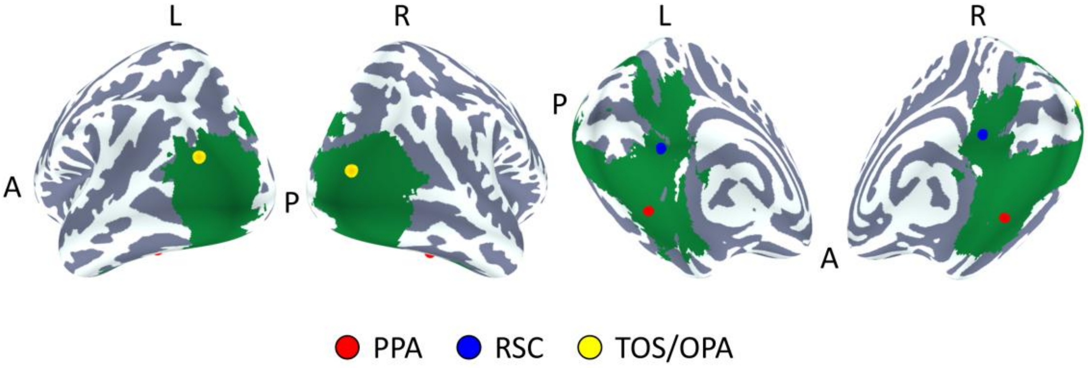


Figure A.1. Mask used for ROI analyses given by the group level contrast of mixed > scrambled.

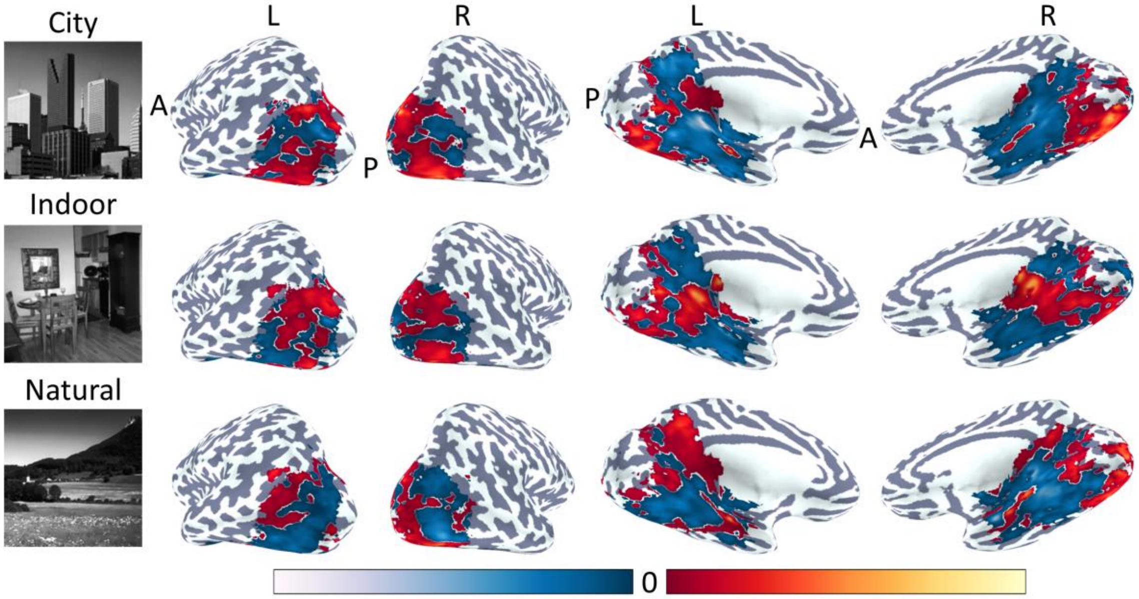


Figure A.2. Patterns of response in Experiment 1 to city, indoor, and natural conditions in a representative participant. Patterns are restricted to regions defined by the response of mixed scenes > scrambled scenes. Red and blue colours indicate values above and below the mean respectively.

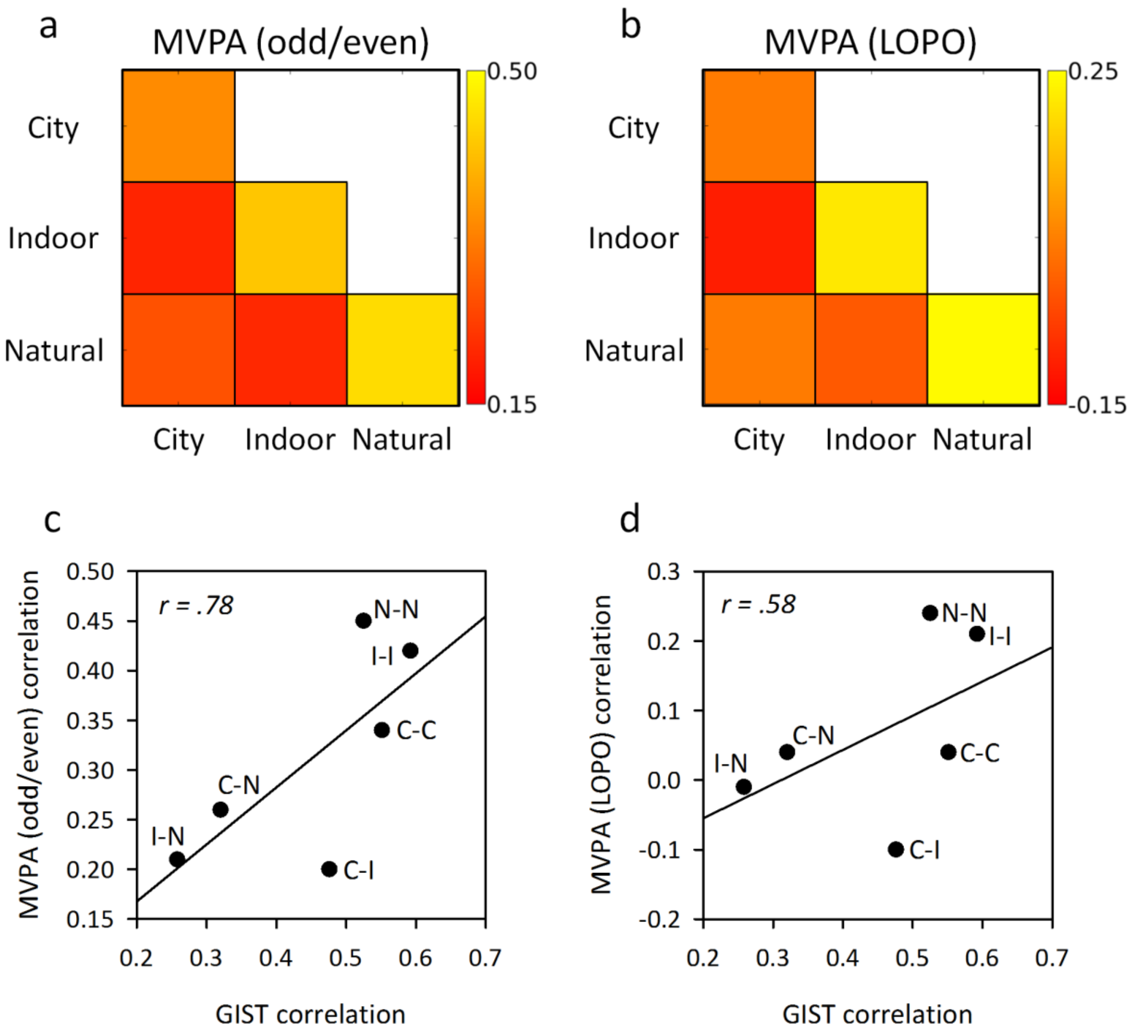


Figure A.3. Experiment 1: Relationship between fMRI responses in ROI restricted to standard scene-selective regions (PPA, RSC, TOS / OPA) and low-level image properties. Within- and between- category correlations for city, indoor, and natural conditions as determined by the individual-participant (a) and LOPO (b) MVPA analyses. Scatter-plots (d-e) showing strong positive correlations of the correlation matrices in (a) and (b) with (Fig. 3.5c) respectively.

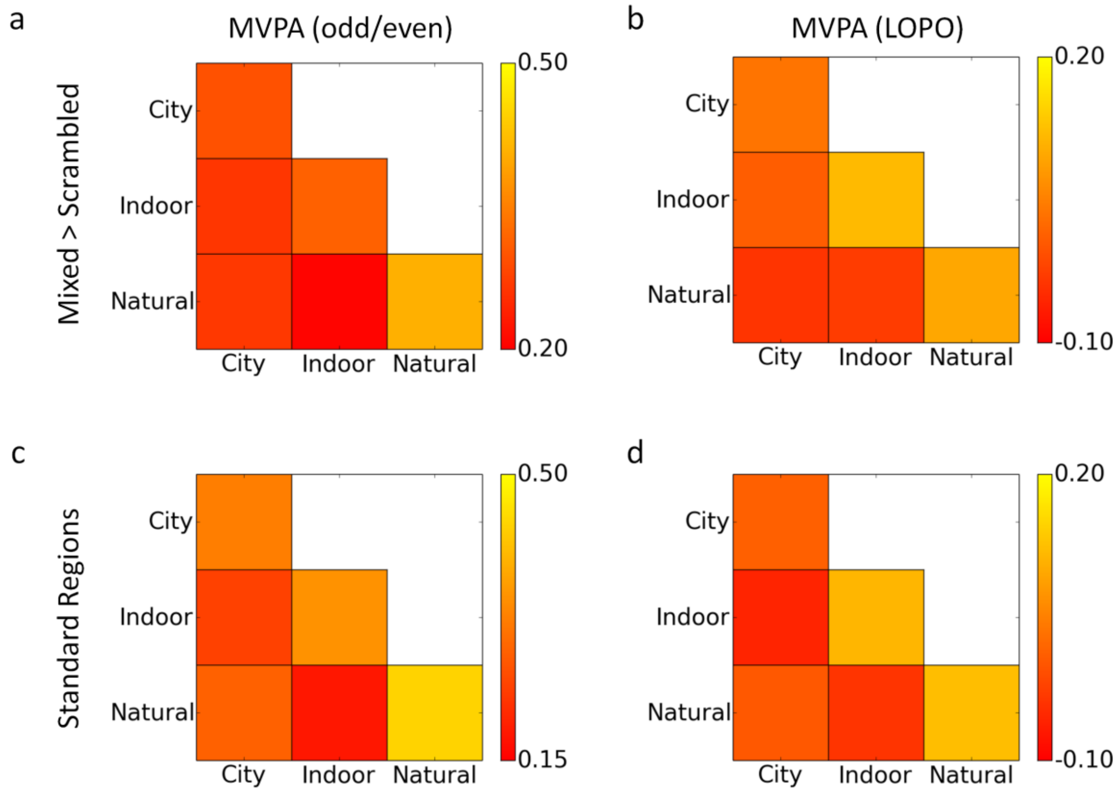


Figure A.4. Experiment 1: MVPA of unsmoothed fMRI data. (a) IP analysis of main scene region (kNN accuracy = 50.0%, t = 3.63, p = .001). (b) LOPO analysis of main scene region (kNN accuracy = 70.83%, t = 8.72, p < .001). (c) IP analysis of standard scene regions (PPA, RSC, TOS/OPA) (kNN accuracy = 52.5%, t = 4.80, p < .001). (d) LOPO analysis of standard scene regions (kNN accuracy = 62.5%, t = 5.34, p < .001).

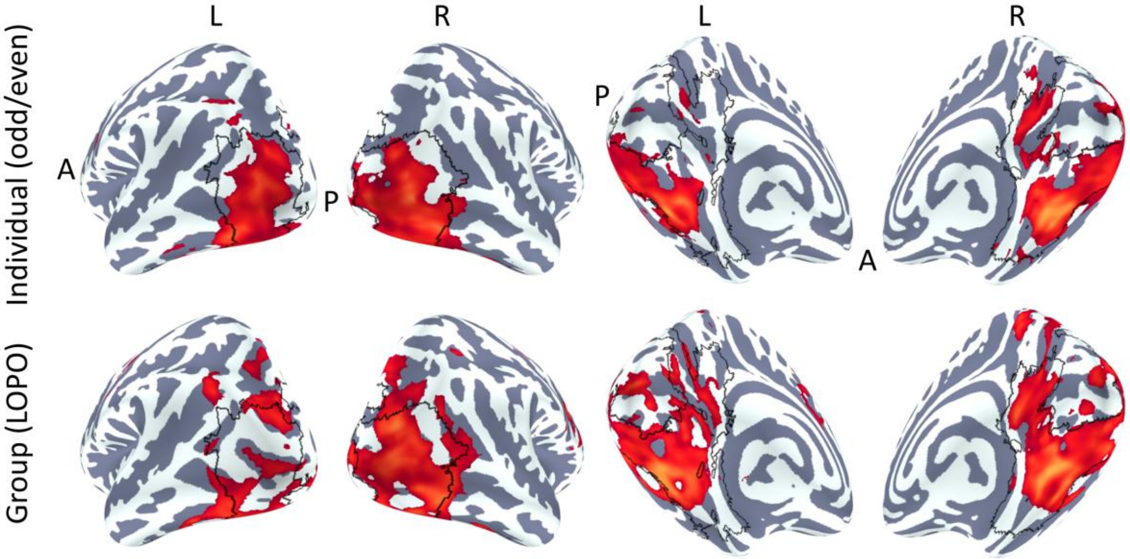


Figure A.5. *Experiment 1: Group statistical maps of searchlight analyses using individual-participant and LOPO paradigms. Thresholded Z>2.3, cluster corrected p<.05. Black border indicates the area of the mixed > scrambled mask used for ROI analyses.*

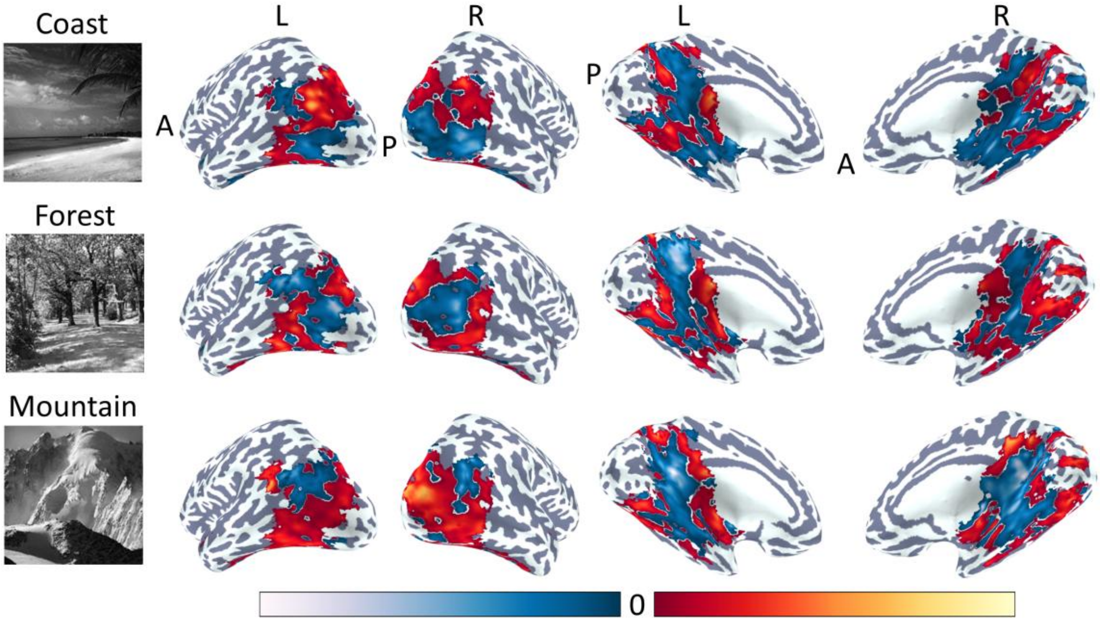


Figure A.6. Patterns of response in Experiment 2 to coast, forest, and mountain conditions in a representative participant. Patterns are restricted to regions defined by the response of mixed scenes > scrambled scenes. Red and blue colours indicate values above and below the mean respectively.

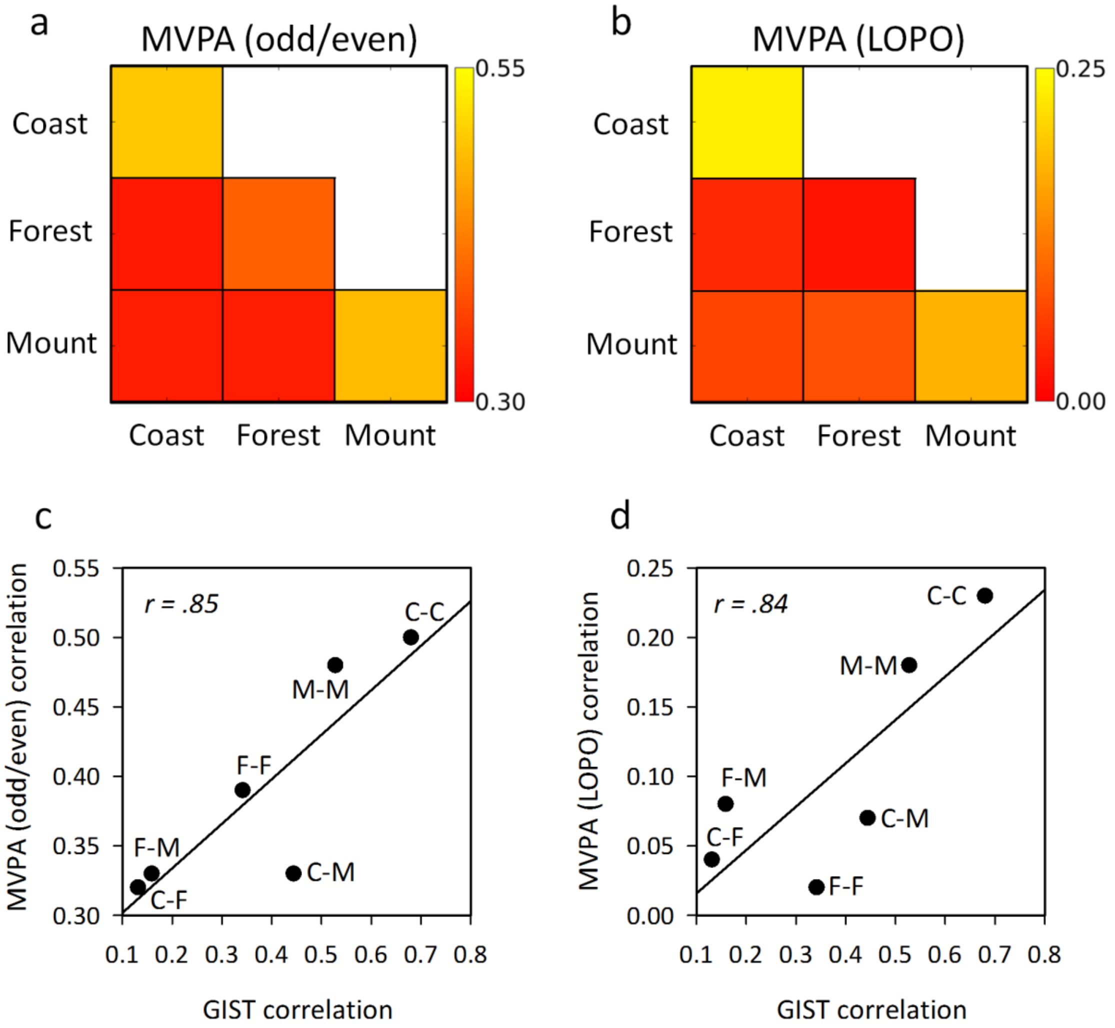


Figure A.7. Experiment 2: Relationship between fMRI responses in ROI restricted to standard scene-selective regions (PPA, RSC, TOS / OPA) and low-level image properties. Within- and between- category correlations for coast, forest, and mountain conditions as determined by the individual-participant (a) and LOPO (b) MVPA analyses. Scatter-plots (d-e) showing strong positive correlations of the correlation matrices in (a) and (b) with (Fig. 3.7c) respectively.

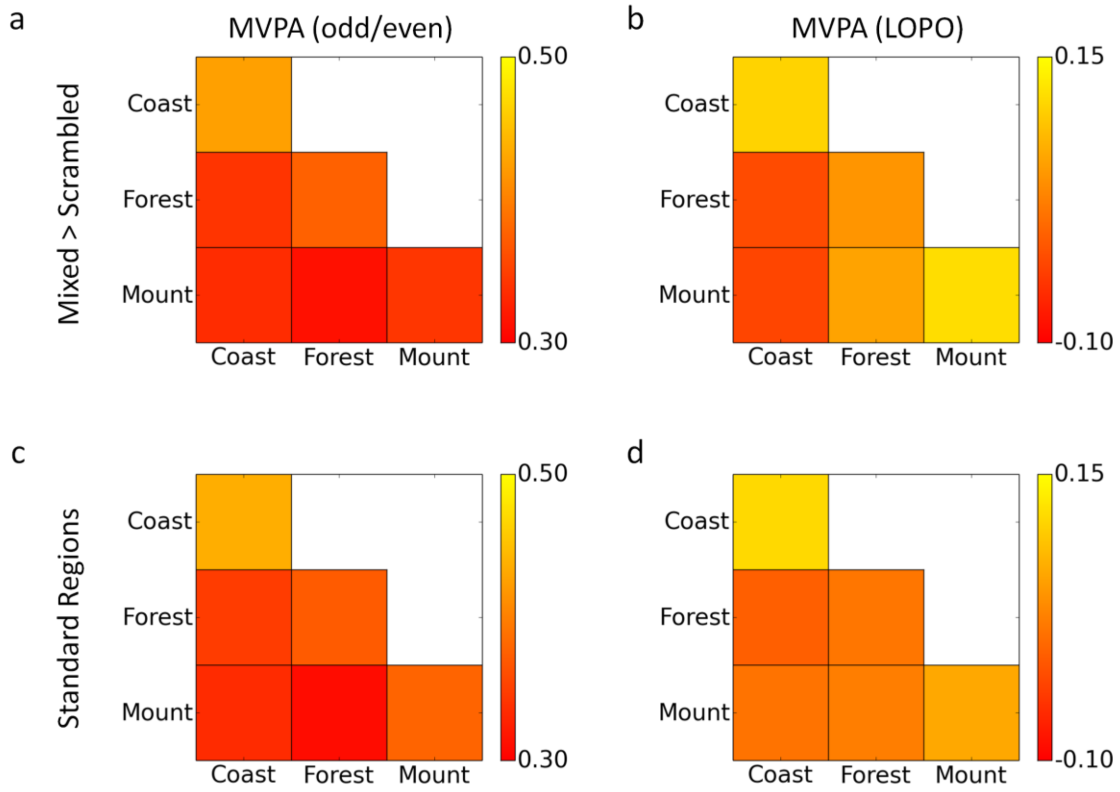


Figure A.8. Experiment 2: MVPA of unsmoothed fMRI data. (a) IP analysis of main scene region (kNN accuracy = 45.8%, t = 2.66, p = .015). (b) LOPO analysis of main scene region (kNN accuracy = 70.0%, t = 7.75, p < .001). (c) IP analysis of standard scene regions (PPA, RSC, TOS/OPA) (kNN accuracy = 45.8%, t = 3.08, p = .006). (d) LOPO analysis of standard scene regions (kNN accuracy = 56.7%, t = 5.80, p < .001).

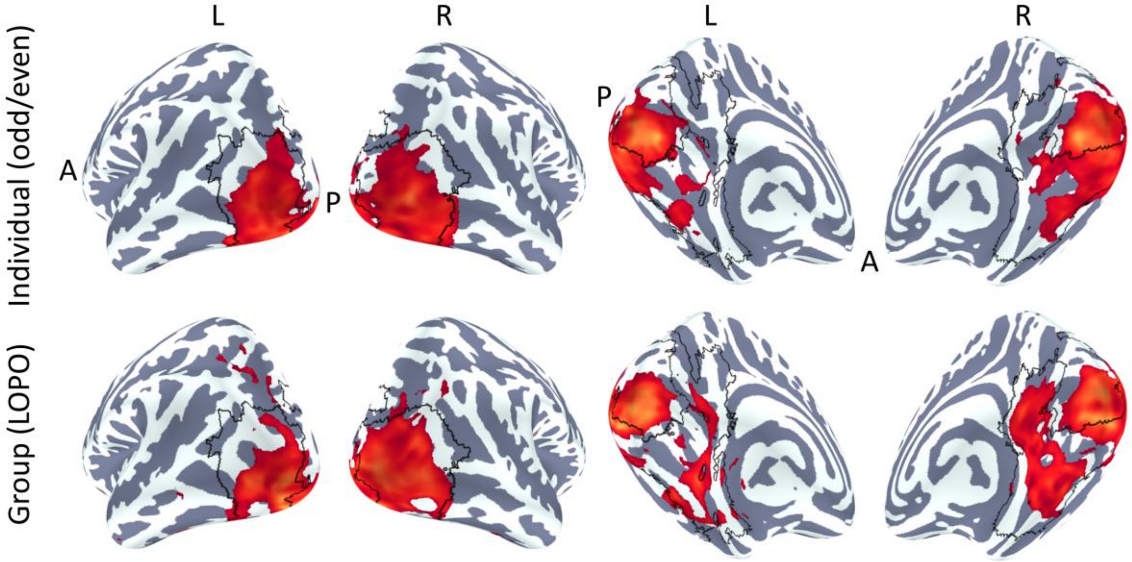


Figure A.9. Experiment 2: Group statistical maps of searchlight analyses using individual-participant and LOPO paradigms. Thresholded Z>2.3, cluster corrected p<.05. Black border indicates the area of the mixed > scrambled mask used for ROI analyses.

A.1.2 Chapter 4

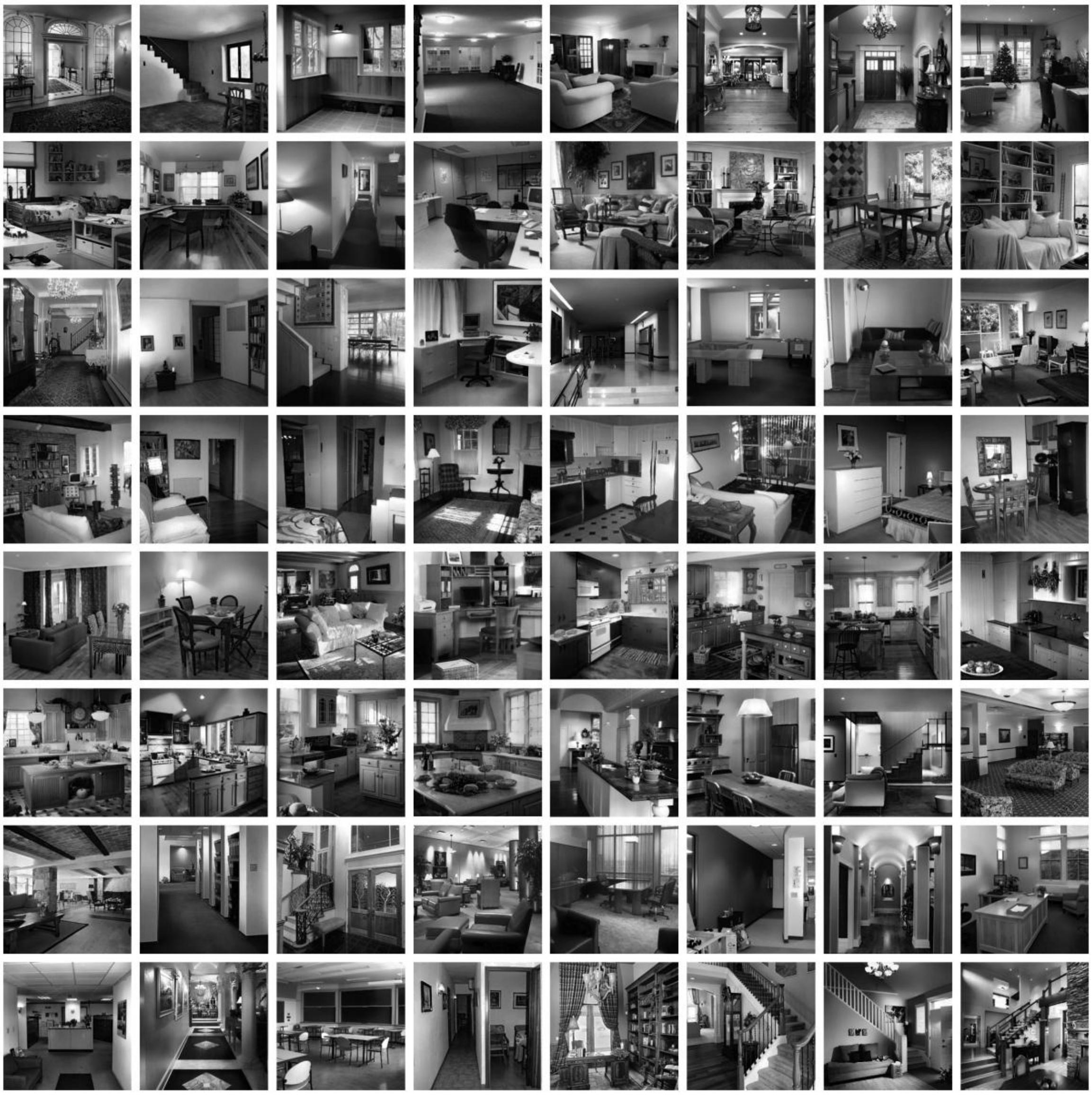


Figure A.10. Full unfiltered indoor scene image set.

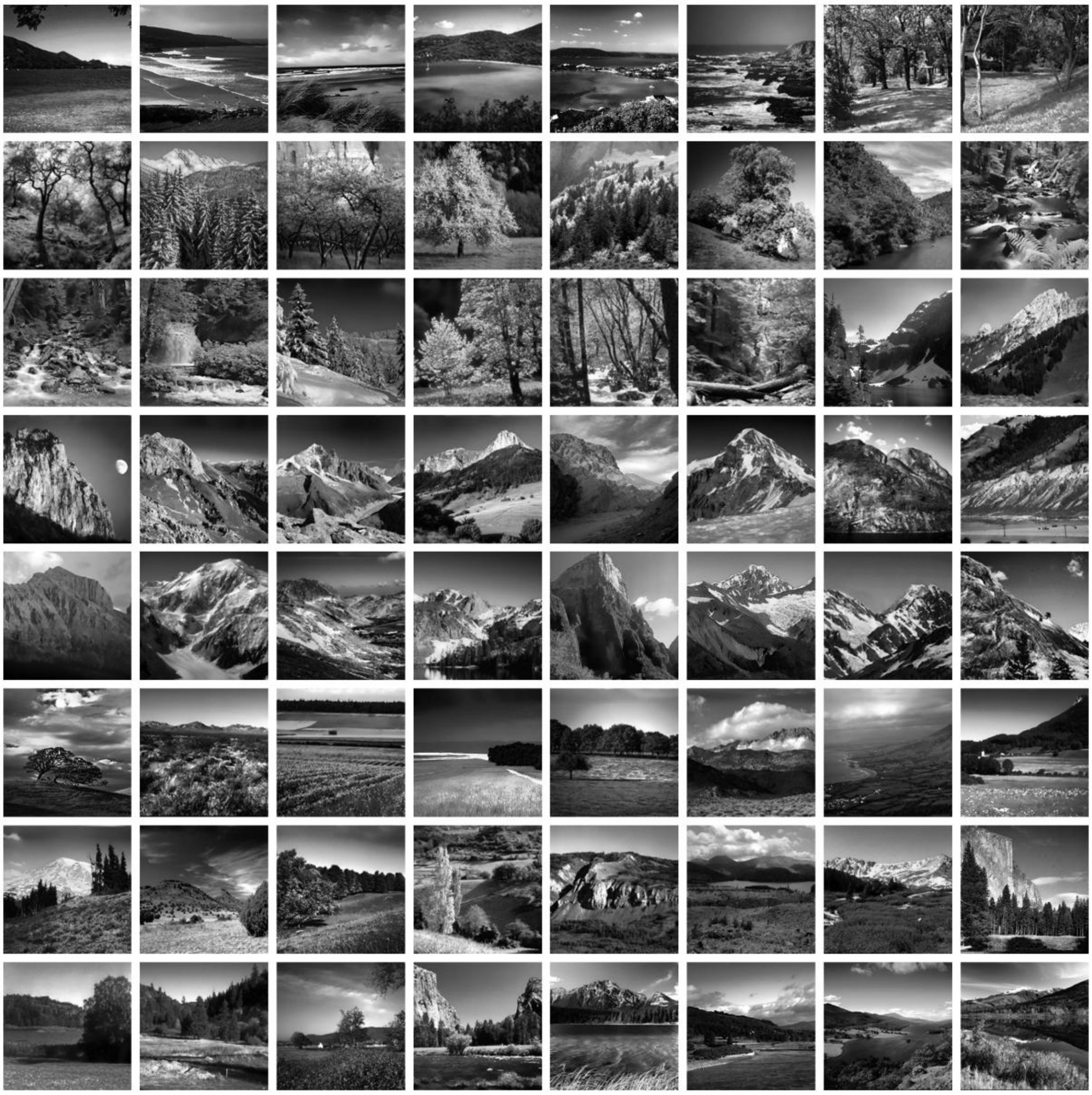


Figure A.11. Full unfiltered natural scene image set.

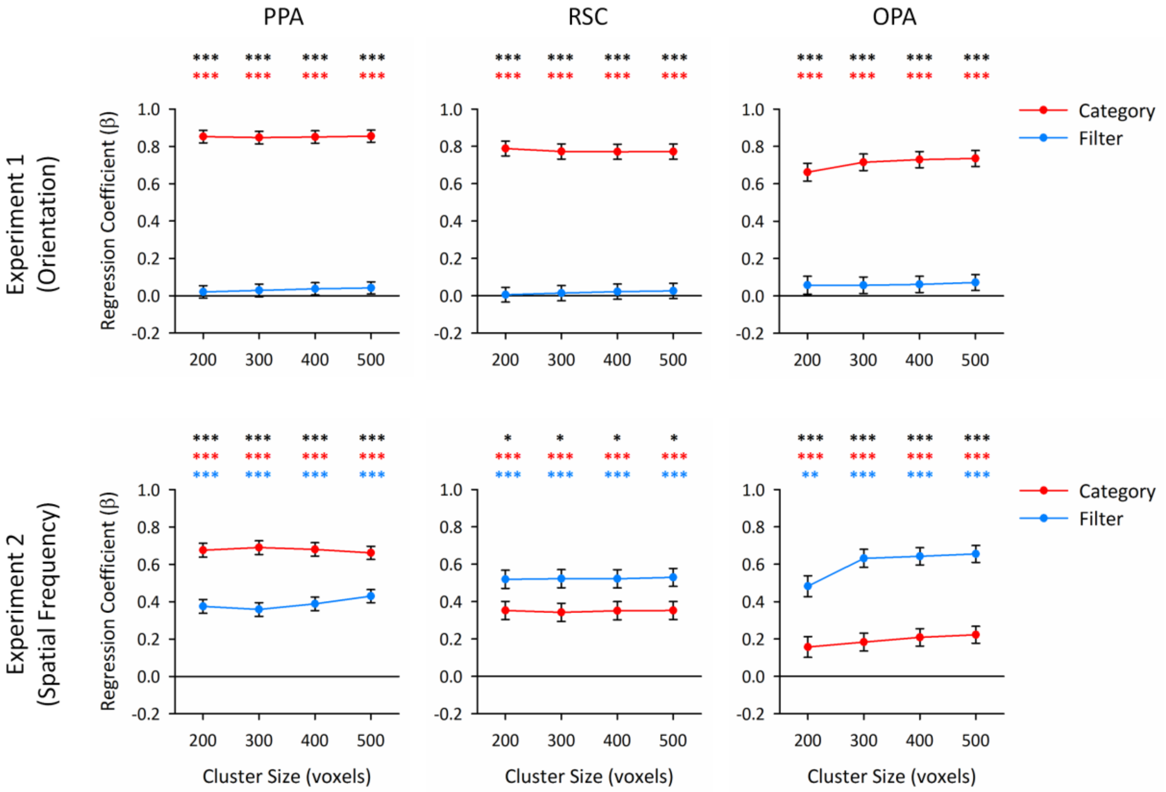


Figure A.12. A flood-fill algorithm was used to identify ROIs for each of the scene-selective regions (PPA, RSC, OPA) in each hemisphere. Clusters were defined to comprise 200, 300, 400, or 500 contiguous voxels, and then combined across hemispheres for each region to yield final ROIs comprising 400, 600, 800, or 1000 voxels respectively. The multi-voxel pattern analyses and representational similarity analyses were conducted for each ROI independently. The resulting regression coefficients are displayed above; coloured asterisks indicate the significance of the corresponding regressors, whilst black asterisks indicate the significance of the contrast between the regressors (*** p < .001, ** p < .01, * p < .05). Error bars represent 1 SEM. In all cases, cluster size is seen to have little effect upon the results of the regression analyses.

A.1.3 Chapter 5

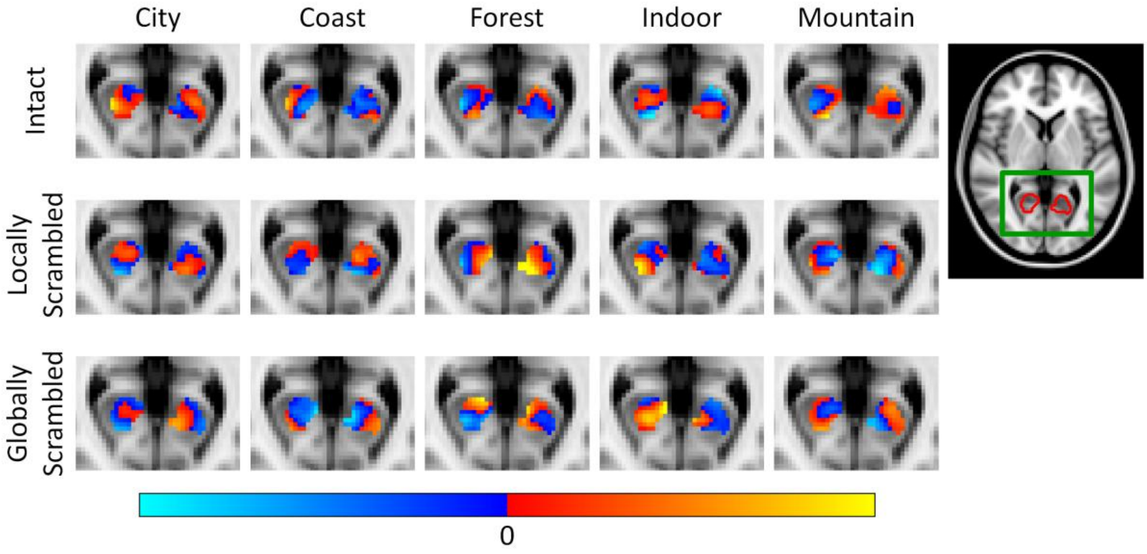


Figure A.13. Group patterns of response for each condition, restricted to RSC region. Responses within each level of scrambling are normalized by subtracting a voxel-wise mean across all categories, such that red and blue colours indicate values above and below the mean respectively.

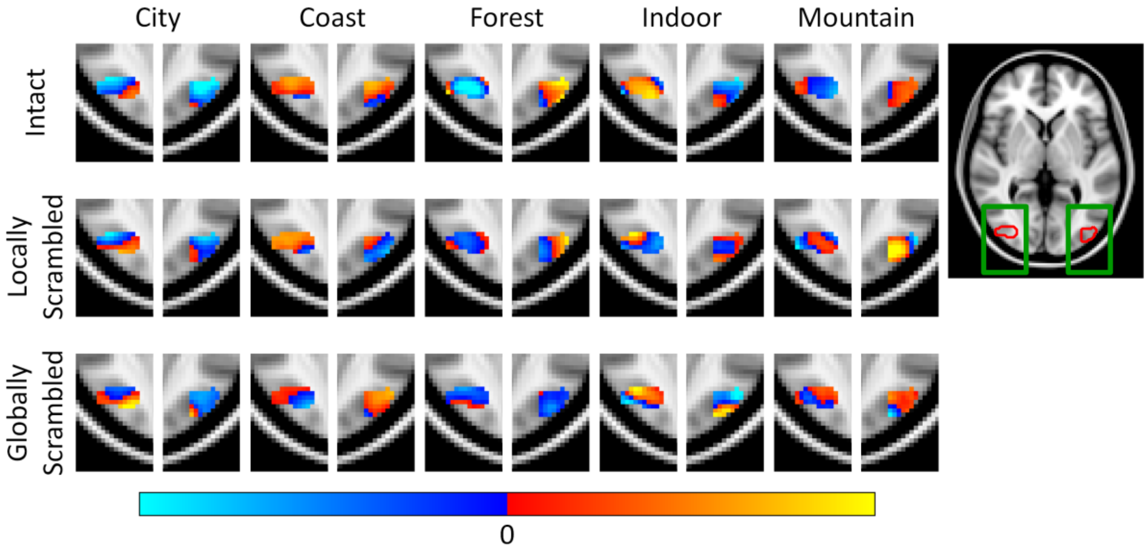


Figure A.14. Group patterns of response for each condition, restricted to OPA region. Responses within each level of scrambling are normalized by subtracting a voxel-wise mean across all categories, such that red and blue colours indicate values above and below the mean respectively.

A.2 Supplementary Tables

A.2.1 Chapter 3

Table A.1. MNI mm co-ordinates (x, y, z) of PPA, RSC, and TOS / OPA regions reported in literature.

		LH	RH
PPA	Dilks et al. (2011)	-25, -45, -6	27, -45, -8
	Epstein et al. (1999)	-29, -40, -7	23, -40, -7
	Epstein et al. (2003)	-27, -51, -9	31, -48, -12
	Epstein and Higgins (2007)	-19, -37, -8	20, -36, -6
	Golomb and Kanwisher (2011)	-28, -52, -10	28, -51, -12
	Henderson et al. (2011)	-19, -42, -2	23, -41, -3
	Köhler et al. (2002)	-12, -42, -2	21, -35, -11
	Mullally and Maguire (2011)	-27, -42, -12	33, -39, -12
	O'Craven and Kanwisher (2000)	-28, -39, -3	31, -39, -6
	Park et al. (2007)	-26, -42, -12	26, -42, -11
RSC	Dilks et al. (2011)	-19, -57, 15	21, -56, 6
	Epstein and Higgins (2007)	-10, -59, 8	13, -54, 9
	Park et al. (2007)	-16, -55, 20	15, -51, 22
	Schinazi and Epstein (2010)	-22, -50, 6	17, -53, 12
TOS / OPA	Dilks et al. (2011)	-34, -78, 27	38, -75, 26
	Epstein and Higgins (2007)	-33, -79, 31	32, -75, 34
	Hasson et al. (2003)	-35, -81, 18	37, -79, 16
	Levy et al. (2004)	-36, -80, 17	36, -78, 19

Bibliography

Aguirre, G. K., & D'Esposito, M. (1997). Environmental knowledge is subserved by separable dorsal/ventral neural areas. *Journal of Neuroscience*, *17*(7), 2512–2518.

Aguirre, G. K., & D'Esposito, M. (1999). Topographical disorientation: a synthesis and taxonomy. *Brain*, *122*(9), 1613–1628.

Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). An area within human ventral cortex sensitive to "building" stimuli: evidence and implications. *Neuron*, *21*(2), 373–83.

Aminoff, E., Gronau, N., & Bar, M. (2007). The parahippocampal cortex mediates spatial and nonspatial associations. *Cerebral Cortex*, *17*(7), 1493–503.

Amit, E., Mehoudar, E., Trope, Y., & Yovel, G. (2012). Do object-category selective regions in the ventral visual stream represent perceived distance information? *Brain and Cognition*, *80*(2), 201–213.

Amunts, K., Malikovic, A., Mohlberg, H., Schormann, T., & Zilles, K. (2000). Brodmann's areas 17 and 18 brought into stereotaxic space - where and how variable? *NeuroImage*, *11*(1), 66–84.

Andrews, T. J., Watson, D. M., Rice, G. E., & Hartley, T. (2015). Low-level properties of natural images predict topographic patterns of neural response in the ventral visual pathway visual pathway. *Journal of Vision*, *15*(7), 1–12.

Arcaro, M. J., McMains, S. A., Singer, B. D., & Kastner, S. (2009). Retinotopic Organization of Human Ventral Visual Cortex. *Journal of Neuroscience*, *29*(34), 10638–10652.

Baldassano, C., Beck, D. M., & Fei-Fei, L. (2013). Differential connectivity within the Parahippocampal Place Area. *NeuroImage*, *75*, 228–37.

Bar, M., Aminoff, E., & Ishai, A. (2008a). Famous Faces Activate Contextual Associations in the Parahippocampal Cortex. *Cerebral Cortex*, *18*(6), 1233–1238.

Bar, M., Aminoff, E., & Schacter, D. L. (2008b). Scenes Unseen: The Parahippocampal

Cortex Intrinsically Subserves Contextual Associations, Not Scenes or Places Per Se. *Journal of Neuroscience*, *28*(34), 8539–8544.

Barrash, J., Damasio, H., Adolphs, R., & Tranel, D. (2000). The neuroanatomical correlates of route learning impairment. *Neuropsychologia*, *38*(6), 820–836.

Burgess, N., Becker, S., King, J. A., & O'Keefe, J. (2001). Memory for events and their spatial context: models and experiments. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *356*(1413), 1493–1503.

Byrne, P., Becker, S., & Burgess, N. (2007). Remembering the past and imagining the future: A neural model of spatial memory and imagery. *Psychological Review*, *114*(2), 340–375.

Carlson, T., Tovar, D. A., & Kriegeskorte, N. (2013). Representational dynamics of object vision : The first 1000 ms. *Journal of Vision*, *13*(10), 1–19.

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature neuroscience*, *17*(3), 455–462.

Cichy, R. M., Sterzer, P., Heinzle, J., Elliott, L. T., Ramirez, F., & Haynes, J.-D. (2013). Probing principles of large-scale object representation: Category preference and location encoding. *Human Brain Mapping*, *34*(7), 1636–1651.

Clarke, A., Devereux, B. J., Randall, B., & Tyler, L. K. (2015). Predicting the Time Course of Individual Objects with MEG. *Cerebral Cortex*, *25*(October), 3602–3612.

Clithero, J. A., Smith, D. V, Carter, R. M., & Huettel, S. A. (2011). Within- and cross-participant classifiers reveal different neural coding of information. *Neuroimage*, *56*(2), 699–708.

Dalal, N., & Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*.

Davatzikos, C., Ruparel, K., Fan, Y., Shen, D. G., Acharyya, M., Loughead, J. W., … Langleben, D. D. (2005). Classifying spatial patterns of brain activity with machine learning methods: Application to lie detection. *Neuroimage*, *28*(3), 663–668.

Dilks, D. D., Julian, J. B., Kubilius, J., Spelke, E. S., & Kanwisher, N. (2011). Mirror-Image Sensitivity and Invariance in Object and Scene Processing Pathways. *Journal of Neuroscience*, *31*(31), 11305–11312.

Dilks, D. D., Julian, J. B., Paunov, A. M., & Kanwisher, N. (2013). The Occipital Place Area Is Causally and Selectively Involved in Scene Perception. *Journal of Neuroscience*, *33*(4), 1331–1336.

Ehinger, K. A., Xiao, J. X., Torralba, A., & Oliva, A. (2011). Estimating scene typicality from human ratings and image features. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2562–2567). Austin, Tx: Cognitive Science Society.

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage*, *25*(4), 1325–35.

Epstein, R. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends in Cognitive Sciences*, *12*(10), 388–396.

Epstein, R. A., Higgins, J. S., Parker, W., Aguirre, G. K., & Cooperman, S. (2006). Cortical correlates of face and scene inversion: A comparison. *Neuropsychologia*, *44*(7), 1145–1158.

Epstein, R. A., Higgins, J. S., & Thompson-Schill, S. L. (2005). Learning places from views: Variation in scene processing as a function of experience and navigational ability. *Journal of Cognitive Neuroscience*, *17*(1), 73–83.

Epstein, R. A., Parker, W. E., & Feiler, A. M. (2007a). Where am I now? Distinct roles for parahippocampal and retrosplenial cortices in place recognition. *Journal of Neuroscience*, *27*(23), 6141–6149.

Epstein, R. A., & Ward, E. J. (2010). How Reliable Are Visual Context Effects in the Parahippocampal Place Area? *Cerebral Cortex*, *20*(2), 294–303.

Epstein, R., Graham, K. S., & Downing, P. E. (2003). Viewpoint-specific scene representations in human parahippocampal cortex. *Neuron*, *37*(5), 865–876.

Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The parahippocampal place area: Recognition, navigation, or encoding? *Neuron*, *23*(1), 115–125.

Epstein, R., & Higgins, J. S. (2007). Differential parahippocampal and retrosplenial involvement in three types of visual scene recognition. *Cerebral Cortex*, *17*(7), 1680–1693.

Epstein, R., Higgins, J. S., Jablonski, K., & Feiler, A. M. (2007b). Visual scene processing in familiar and unfamiliar environments. *Journal of Neurophysiology*, *97*(5), 3670–3683.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*(6676), 598–601.

Etzel, J. A., Zacks, J. M., & Braver, T. S. (2013). Searchlight analysis: promise, pitfalls, and potential. *NeuroImage*, *78*, 261–269.

Ewbank, M. P., Schluppeck, D., & Andrews, T. J. (2005). fMR-adaptation reveals a distributed representation of inanimate objects and places in human visual cortex. *NeuroImage*, *28*(1), 268–79.

Freeman, J., Brouwer, G. J., Heeger, D. J., & Merriam, E. P. (2011). Orientation decoding depends on maps, not columns. *The Journal of Neuroscience*, *31*(13), 4792–4804.

Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C., Frackowiak, R. S., & Turner, R. (1995). Analysis of fMRI time-series revisited. *NeuroImage*, *2*(1), 45–53.

Ganaden, R. E., Mullin, C. R., & Steeves, J. K. E. (2013). Transcranial Magnetic Stimulation to the Transverse Occipital Sulcus Affects Scene but Not Object Processing. *Journal of Cognitive Neuroscience*, *25*(6), 961–968.

Golomb, J. D., & Kanwisher, N. (2012). Higher Level Visual Cortex Represents Retinotopic, Not Spatiotopic, Object Location. *Cerebral Cortex*, *22*(12), 2794–2810.

Gosselin, F., & Schyns, P. G. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Research*, *41*(17), 2261–2271.

Greene, M. R., & Oliva, A. (2006). Natural Scene Categorization from Conjunctions of Ecological Global Properties. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 291–296). Vancouver.

Greene, M. R., & Oliva, A. (2009a). The Briefest of Glances: The Time Course of Natural Scene Understanding. *Psychological Science*, *20*(4), 464–472.

Greene, M. R., & Oliva, A. (2009b). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology*, *58*(2), 137–176.

Greene, M. R., & Oliva, A. (2010). High-Level Aftereffects to Global Scene Properties. *Journal of Experimental Psychology-Human Perception and Performance*, *36*(6), 1430–1442.

Greicius, M. D., Supekar, K., Menon, V., & Dougherty, R. F. (2009). Resting-State Functional Connectivity Reflects Structural Connectivity in the Default Mode Network. *Cerebral Cortex*, *19*(1), 72–78.

Grill-Spector, K. (2003). The neural basis of object perception. *Current Opinion in Neurobiology*, *13*(2), 159–166.

Hanke, M., Halchenko, Y. O., Sederberg, P. B., Hanson, S. J., Haxby, J. V, & Pollmann, S. (2009). PyMVPA: a Python Toolbox for Multivariate Pattern Analysis of fMRI Data. *Neuroinformatics*, *7*(1), 37–53.

Hanson, S. J., Matsuka, T., & Haxby, J. V. (2004). Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: is there a “face” area? *Neuroimage*, *23*(1), 156–166.

Harel, A., Kravitz, D. J., & Baker, C. I. (2014). Task context impacts visual object processing differentially across the cortex. *Proceedings of the National Academy of Sciences*, E962–E971.

Hartley, T., Lever, C., Burgess, N., & O’Keefe, J. (2014). Space in the brain: how the hippocampal formation supports spatial cognition. *Philosophical transactions of the*

Royal Society of London. Series B, Biological sciences, *369*(1635), 20120510.

Hasson, U., Harel, M., Levy, I., & Malach, R. (2003). Large-scale mirror-symmetry organization of human occipito-temporal object areas. *Neuron*, *37*(6), 1027–1041.

Hasson, U., Levy, I., Behrmann, M., Hendler, T., & Malach, R. (2002). Eccentricity bias as an organizing principle for human high-order object areas. *Neuron*, *34*(3), 479–490.

Haxby, J. V, Connolly, A. C., & Guntupalli, J. S. (2014). Decoding Neural Representational Spaces Using Multivariate Pattern Analysis. *Annual Review of Neuroscience*, *37*, 435–456.

Haxby, J. V, Gobbini, M., Furey, M., Ishai, A., Schouten, J., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*(5539), 2425–2430.

Haxby, J. V, Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., … Ramadge, P. J. (2011). A Common, High-Dimensional Model of the Representational Space in Human Ventral Temporal Cortex. *Neuron*, *72*(2), 404–416.

Haxby, J. V, Hoffman, E. A., & Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biological Psychiatry*, *51*(1), 59–67.

Haynes, J.-D. (2015). A Primer on Pattern-Based Approaches to fMRI: Principles, Pitfalls, and Perspectives. *Neuron*, *87*(2), 257–270.

Haynes, J.-D., & Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature Neuroscience*, *8*(5), 686–691.

Henderson, J. M., Zhu, D. C., & Larson, C. L. (2011). Functions of parahippocampal place area and retrosplenial cortex in real-world scene analysis: An fMRI study. *Visual Cognition*, *19*(7), 910–927.

Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A Continuous Semantic Space Describes the Representation of Thousands of Object and Action Categories across the Human Brain. *Neuron*, *76*(6), 1210–1224.

Janzen, G., & Jansen, C. (2010). A neural wayfinding mechanism adjusts for ambiguous landmark information. *Neuroimage*, *52*(1), 364–370.

Janzen, G., Wagensveld, B., & van Turennout, M. (2007). Neural Representation of Navigational Relevance Is Rapidly Induced and Long Lasting. *Cerebral Cortex*, *17*(4), 975–981.

Janzen, G., & Weststeijn, C. G. (2007). Neural representation of object location and route direction: An event-related fMRI study. *Brain Research*, *1165*(0), 116–125.

Jenkins, R., White, D., Van Montfort, X., & Burton, M. A. (2011). Variability in photos of the same face. *Cognition*, *121*(3), 313–323.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage*, *17*(2), 825–841.

Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision Research*, *47*(26), 3286–3297.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature neuroscience*, *8*(5), 679–85.

Kaplan, J. T., & Meyer, K. (2012). Multivariate pattern analysis reveals common neural patterns across individuals during touch observation. *Neuroimage*, *60*(1), 204–212.

Kauffmann, L., Bourgin, J., Guyader, N., & Peyrin, C. (2015a). The Neural Bases of the Semantic Interference of Spatial Frequency-based Information in Scenes. *Journal of Cognitive Neuroscience*, *27*(12), 2394–2405.

Kauffmann, L., Chauvin, A., Guyader, N., & Peyrin, C. (2015b). Rapid scene categorization: Role of spatial frequency order, accumulation mode and luminance contrast. *Vision Research*, *107*, 49–57.

Kauffmann, L., Chauvin, A., Pichat, C., & Peyrin, C. (2015c). Effective connectivity in the neural network underlying coarse-to-fine categorization of visual scenes. A dynamic causal modeling study. *Brain and Cognition*, *99*, 46–56.

Kauffmann, L., Ramanoël, S., Guyader, N., Chauvin, A., & Peyrin, C. (2015d). Spatial frequency processing in scene-selective cortical regions. *NeuroImage*, *112*, 86–95.

Kauffmann, L., Ramanoël, S., & Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. *Frontiers in Intergrative Neuroscience*, *8*(37), 1–14.

Kay, K. N., Weiner, K. S., & Grill-Spector, K. (2015). Attention Reduces Spatial Uncertainty in Human Ventral Temporal Cortex. *Current Biology*, *25*(5), 595–600.

Kim, J. G., Aminoff, E. M., Kastner, S., & Behrmann, X. M. (2015). A Neural Basis for Developmental Topographic Disorientation. *The Journal of Neuroscience*, *35*(37), 12954–12969.

Kim, M., Ducros, M., Carlson, T., Ronen, I., He, S., Ugurbil, K., & Kim, D. S. (2006). Anatomical correlates of the functional organization in the human occipitotemporal cortex. *Magnetic Resonance Imaging*, *24*(5), 583–590.

Kobayashi, Y., & Amaral, D. G. (2007). Macaque monkey retrosplenial cortex: III. Cortical efferents. *The Journal of Comparative Neurology*, *502*(5), 810–833.

Köhler, S., Crane, J., & Milner, B. (2002). Differential contributions of the parahippocampal place area and the anterior hippocampus to human memory for scenes. *Hippocampus*, *12*(6), 718–723.

Kravitz, D. J., Peng, C. S., & Baker, C. I. (2011). Real-World Scene Representations in High-Level Visual Cortex: It's the Spaces More Than the Places. *Journal of Neuroscience*, *31*(20), 7322–7333.

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences*, *103*(10), 3863–3868.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*(4), 1–28.

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., & Baker, C. I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nature Neuroscience*, *12*(5), 535–540.

Lescroart, M. D., Stansbury, D. E., & Gallant, J. L. (2015). Fourier power, subjective distance, and object categories all provide plausible models of BOLD responses in scene-selective visual areas. *Frontiers in Computational Neuroscience*, *9*.

Levy, I., Hasson, U., Avidan, G., Hendler, T., & Malach, R. (2001). Center – periphery organization of human object areas. *Nature Neuroscience*, *4*(5), 533–539.

Levy, I., Hasson, U., Harel, M., & Malach, R. (2004). Functional analysis of the periphery effect in human building-related areas. *Human Brain Mapping*, *22*(1), 15–26.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*(2), 91–110.

MacEvoy, S. P., & Epstein, R. A. (2007). Position selectivity in scene- and object-responsive occipitotemporal regions. *Journal of Neurophysiology*, *98*(4), 2089–2098.

Maguire, E. (2001). The retrosplenial contribution to human navigation: A review of lesion and neuroimaging findings. *Scandinavian Journal of Psychology*, *42*(3), 225–238.

Malach, R., Levy, I., & Hasson, U. (2002). The topography of high-order human object areas. *Trends in Cognitive Sciences*, *6*(4), 176–184.

Marchette, S. A., Vass, L. K., Ryan, J., & Epstein, R. A. (2014). Anchoring the neural compass: coding of local spatial reference frames in human medial parietal lobe. *Nature Neuroscience*, *17*, 1598–1606.

Marchette, S. A., Vass, L. K., Ryan, J., & Epstein, R. A. (2015). Outside Looking In: Landmark Generalization in the Human Navigational System. *Journal of Neuroscience*, *35*(44), 14896–14908.

McCotter, M., Gosselin, F., Sowden, P., & Schyns, P. (2005). The use of visual information in natural scenes. *Visual Cognition*, *12*(6), 938–953.

Mendez, M. F., & Cherrier, M. M. (2003). Agnosia for scenes in topographagnosia. *Neuropsychologia*, *41*(10), 1387–1395.

Misaki, M., Kim, Y., Bandettini, P. A., & Kriegeskorte, N. (2010). Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *Neuroimage*, *53*(1), 103–118.

Mourão-Miranda, J., Bokde, A. L. W., Born, C., Hampel, H., & Stetter, M. (2005). Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on functional MRI data. *NeuroImage*, *28*(4), 980–95.

Mullally, S. L., & Maguire, E. A. (2011). A new role for the parahippocampal cortex in representing space. *The Journal of Neuroscience*, *31*(20), 7441–9.

Mullin, C. R., & Steeves, J. K. E. (2013). Consecutive TMS-fMRI Reveals an Inverse Relationship in BOLD Signal between Object and Scene Processing. *Journal of Neuroscience*, *33*(49), 19243–19249.

Mur, M., Bandettini, P. A., & Kriegeskorte, N. (2009). Revealing representational content with pattern-information fMRI - an introductory guide. *Social Cognitive and Affective Neuroscience*, *4*(1), 101–109.

Musel, B., Kauffmann, L., Ramanoël, S., Giavarini, C., Guyader, N., Chauvin, A., & Peyrin, C. (2014). Coarse-to-fine Categorization of Visual Scenes in Scene-selective Cortex. *Journal of Cognitive Neuroscience*, *26*(10), 2287–2297.

Naselaris, T., Prenger, R. J., Kay, K. N., Oliver, M., & Gallant, J. L. (2009). Bayesian Reconstruction of Natural Images from Human Brain Activity. *Neuron*, *63*(6), 902–915.

Nasr, S., Echavarria, C. E., & Tootell, R. B. H. (2014). Thinking Outside the Box: Rectilinear Shapes Selectively Activate Scene-Selective Cortex. *Journal of Neuroscience*, *34*(20), 6721–6735.

Nasr, S., Liu, N., Devaney, K. J., Yue, X., Rajimehr, R., Ungerleider, L. G., & Tootell, R. B. H. (2011). Scene-Selective Cortical Regions in Human and Nonhuman Primates. *Journal of Neuroscience*, *31*(39), 13771–13785.

Nasr, S., & Tootell, R. B. H. (2012). A cardinal orientation bias in scene-selective visual cortex. *The Journal of Neuroscience*, *32*(43), 14921–6.

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A Toolbox for Representational Similarity Analysis. *PLoS Computational Biology*, *10*(4), e1003553.

O'Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stiimulus-specific brain regions. *Journal of cognitive neuroscience*, *12*(6), 1013–23.

O'Toole, A. J., Jiang, F., Abdi, H., & Haxby, J. V. (2005). Partially distributed representations of objects and faces in ventral temporal cortex. *Journal of Cognitive Neuroscience*, *17*(4), 580–590.

Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, *87*(24), 9868–72.

Oliva, A. (2013). Scene Perception. In *The New Visual Neurosciences* (pp. 725–732).

Oliva, A., & Schyns, P. G. (1997). Coarse Blobs or Fine Edges? Evidence That Information Diagnosticity Changes the Perception of Complex Visual Stimuli. *Cognitive Psychology*, *34*, 72–107.

Oliva, A., & Torralba, A. (2001). Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. *International Journal of Computer Vision*, *42*(3), 145–175.

Oosterhof, N. N., Wiestler, T., Downing, P. E., & Diedrichsen, J. (2011). A comparison of volume-based and surface-based multi-voxel pattern analysis. *NeuroImage*, *56*(2), 593–600.

Op de Beeck, H. P. (2010). Against hyperacuity in brain reading: spatial smoothing does not hurt multivariate fMRI analyses? *NeuroImage*, *49*(3), 1943–8.

Op de Beeck, H. P., Haushofer, J., & Kanwisher, N. G. (2008). Interpreting fMRI data: maps, modules and dimensions. *Nature Reviews Neuroscience*, *9*(2), 123–135.

Park, S., Brady, T. F., Greene, M. R., & Oliva, A. (2011). Disentangling Scene Content from Spatial Boundary: Complementary Roles for the Parahippocampal Place Area and Lateral Occipital Complex in Representing Real-World Scenes. *Journal of Neuroscience*, *31*(4), 1333–1340.

Park, S., & Chun, M. M. (2009). Different roles of the parahippocampal place area (PPA) and retrosplenial cortex (RSC) in panoramic scene perception. *Neuroimage*, *47*(4), 1747–1756.

Park, S., Intraub, H., Yi, D.-J., Widders, D., & Chun, M. M. (2007). Beyond the Edges of a View: Boundary Extension in Human Scene-Selective Visual Cortex. *Neuron*, *54*(2), 335–342.

Park, S., Konkle, T., & Oliva, A. (2015). Parametric Coding of the Size and Clutter of Natural Scenes in the Human Brain. *Cerebral Cortex*, *25*, 1792–1805.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Peirce, J. W. (2007). PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*, 8–13.

Peirce, J. W. (2009). Generating Stimuli for Neuroscience Using PsychoPy. *Frontiers in Neuroinformatics*, *2*(10), 1–8.

Pereira, F., & Botvinick, M. (2011). Information mapping with pattern classifiers: A comparative study. *NeuroImage*, *56*(2), 476–496.

Poldrack, R. A., Halchenko, Y. O., & Hanson, S. J. (2009). Decoding the large-scale structure of brain function by classifying mental states across individuals. *Psychological Science*, *20*(11), 1364–72.

Potter, M. C. (1975). Meaning in visual search. *Science*, *187*(4180), 965–966.

Pugeault, N., & Bowden, R. (2011). Driving me around the bend: Learning to drive from visual gist. In *2011 IEEE International Conference on Computer Vision Workshops*

(ICCV Workshops) (pp. 1022–1029).

Quattoni, A., & Torralba, A. (2009). Recognizing Indoor Scenes. In *Cvpr: 2009 Ieee Conference on Computer Vision and Pattern Recognition, Vols 1-4* (pp. 413–420). New York: Ieee.

Raizada, R. D. S., & Lee, Y. (2013). Smoothness without Smoothing : Why Gaussian Naive Bayes Is Not Naive for Multi-Subject Searchlight Studies. *PLoS ONE*, *8*(7), e69566.

Rajimehr, R., Devaney, K. J., Bilenko, N. Y., Young, J. C., & Tootell, R. B. H. (2011). The "Parahippocampal Place Area" Responds Preferentially to High Spatial Frequencies in Humans and Monkeys. *PLoS Biol*, *9*(4), e1000608.

Rice, G. E., Watson, D. M., Hartley, T., & Andrews, T. J. (2014). Low-level image properties of visual objects predict patterns of neural response across category-selective regions of the ventral visual pathway. *Journal of Neuroscience*, *34*(26), 8837–8844.

Riesenhuber, M., & Poggio, T. L. B. natureneuro. 2. 1999. 101. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*(11), 1019–1025.

Russell, B. C., Torralba, A., Murphy, K. P., & Freeman, W. T. (2008). LabelMe: a database and web-based tool for image annotation. *International Journal of Computer Vision*, *77*(1-3), 157–173.

Schinazi, V. R., & Epstein, R. A. (2010). Neural correlates of real-world route learning. *Neuroimage*, *53*(2), 725–735.

Schyns, P. G., & Oliva, A. (1994). From Blobs to Boundary Edges: Evidence for Time- and Spatial-Scale-Dependent Scene Recognition. *Psychological Science*, *5*(4), 195–200.

Schyns, P. G., & Oliva, A. (1999). Dr. Angry and Mr. Smile: when categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition*, *69*(3), 243–65.

Shinkareva, S. V, Malave, V. L., Mason, R. A., Mitchell, T. M., & Just, M. A. (2011). Commonality of neural representations of words and pictures. *NeuroImage*, *54*(3), 2418–25.

Shinkareva, S. V, Mason, R. A., Malave, V. L., Wang, W., Mitchell, T. M., & Just, M. A. (2008). Using FMRI brain activation to identify cognitive states associated with perception of tools and dwellings. *PloS one*, *3*(1), e1394.

Silson, E. H., Chan, a. W.-Y., Reynolds, R. C., Kravitz, D. J., & Baker, C. I. (2015). A Retinotopic Basis for the Division of High-Level Scene Processing between Lateral and Ventral Human Occipitotemporal Cortex. *Journal of Neuroscience*, *35*(34), 11921–11935.

Stansbury, D. E., Naselaris, T., & Gallant, J. L. (2013). Natural Scene Statistics Account for the Representation of Scene Categories in Human Visual Cortex. *Neuron*, *79*(5), 1025–1034.

Stelzer, J., Chen, Y., & Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): Random permutations and cluster size control. *NeuroImage*, *65*, 69–82.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., … Rabinovich, A. (2015). Going Deeper with Convolutions. *Cvpr*.

Torralba, A. (2009). How many pixels make an image? *Visual Neuroscience*, *26*(1), 123–131.

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network: Computation in Neural Systems*, *14*(3), 391–412.

Torralbo, A., Walther, D. B., Chai, B., Caddigan, E., Fei-Fei, L., & Beck, D. M. (2013). Good Exemplars of Natural Scene Categories Elicit Clearer Patterns than Bad Exemplars but Not Greater BOLD Activity. *PLoS ONE*, *8*(3), e58594.

Vann, S. D., Aggleton, J. P., & Maguire, E. a. (2009). What does the retrosplenial cortex do? *Nature Reviews Neuroscience*, *10*(11), 792–802.

Vass, L. K., & Epstein, R. A. (2013). Abstract representations of location and facing direction in the human brain. *The Journal of Neuroscience*, *33*(14), 6133–42.

Walther, D. B., Caddigan, E., Fei-Fei, L., & Beck, D. M. (2009). Natural Scene Categories

Revealed in Distributed Patterns of Activity in the Human Brain. *Journal of Neuroscience*, *29*(34), 10573–10581.

Walther, D. B., Chai, B., Caddigan, E., Beck, D. M., & Fei-Fei, L. (2011). Simple line drawings suffice for functional MRI decoding of natural scene categories. *Proceedings of the National Academy of Sciences*, *108*(23), 9661–9666.

Wandell, B. A., & Winawer, J. (2011). Imaging retinotopic maps in the human brain. *Vision research*, *51*(7), 718–37.

Wang, L., Mruczek, R. E., Arcaro, M. J., & Kastner, S. (2015). Probabilistic Maps of Visual Topography in Human Cortex. *Cerebral Cortex*, *25*(October), 3911–3931.

Watson, D. M., Hartley, T., & Andrews, T. J. (2014). Patterns of response to visual scenes are linked to the low-level properties of the image. *NeuroImage*, *99*, 402–410.

Watson, D. M., Hymers, M., Hartley, T., & Andrews, T. J. (2016). Patterns of neural response in scene-selective regions of the human brain are affected by low-level manipulations of spatial frequency. *Neuroimage*, *124*, 107–117.

Wegman, J., & Janzen, G. (2011). Neural Encoding of Objects Relevant for Navigation and Resting State Correlations with Navigational Ability. *Journal of Cognitive Neuroscience*, *23*(12), 3841–3854.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods*, *42*(3), 671–684.

Wolbers, T., Klatzky, R. L., Loomis, J. M., Wutte, M. G., & Giudice, N. A. (2011). Modality-Independent Coding of Spatial Layout in the Human Brain. *Current Biology*, *21*(11), 984–989.

Xiao, J. X., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010). SUN Database: Large-scale Scene Recognition from Abbey to Zoo. In *IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3485–3492). Los Alamitos: IEEE Computer Soc.